

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

UNIVERSITÉ DE SHERBROOKE

Faculté de génie

Département de génie électrique et de génie informatique

**Réduction de bruit
dans un signal vocal pour
les bandes étroite et élargie**

Mémoire de maîtrise ès sciences appliquées

Spécialité : génie électrique

Hassan EZZEDDINE

Sherbrooke (Québec), CANADA

Août 2000



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-67264-6

Canada

Dédicace

**Au sud du Liban et à mes amis
qui ont quitté ce monde pour
que je puisse continuer. Je
vous dois beaucoup ...**

SOMMAIRE

Le bruit de l'environnement dégrade de façon significative le signal de parole. Cette dégradation du signal engendre des problèmes ennuyants pour l'oreille humaine et affecte les performances de la plupart des systèmes de communications par voix. À titre d'exemple, le bruit extérieur engendré par des éléments en mouvement (voitures, camions, avions, ...) nuit à la conversation qui se fait par le biais d'un téléphone public.

La solution réside dans la recherche de méthodes permettant de séparer le bruit du signal de parole original. Avec l'arrivée de la téléphonie cellulaire, la téléconférence et la reconnaissance de la parole, le sujet de la réduction du bruit revêt une importance grandissante.

Dans le cadre de ce travail, nous nous sommes efforcés à présenter deux algorithmes qui réduisent de façon significative le bruit additif qui s'ajoute au signal de parole propre.

Le premier algorithme est un réducteur de bruit pour la bande téléphonique de 4 kHz. La base initiale de cet algorithme a été réalisée par Ramabadran et al [11]. Mais, nous l'avons modifié pour le rendre plus efficace parce qu'il modifie un signal de parole propre (sans bruit) et atténue les basses fréquences dans toutes les situations. Cet algorithme se base sur la soustraction spectrale et le rapport signal sur bruit (RSB) par bandes spectrales. Le calcul du RSB a pour but d'éviter la soustraction du bruit dans les bandes où le RSB est élevé. Ainsi, la quantité de bruit à soustraire dépend du niveau du RSB. Cette façon de faire diminue le bruit musical causé par la

soustraction spectrale. Les résultats de cet algorithme ont été comparés avec ceux d'un autre algorithme présenté par Ramabadran et al, et se sont avérés meilleurs.

Après avoir compris, testé et amélioré cet algorithme pour la bande étroite et pour répondre au besoin de la technologie actuelle nous avons étendu cet algorithme pour une bande élargie à 8 kHz.

REMERCIEMENTS

Au début, je remercie Dieu qui m'a donné la santé et qui m'a offert la patience afin d'accomplir ce travail.

Je tiens à remercier mon directeur de recherche, le professeur Dr. Redwan Salami, pour son amitié, sa grande disponibilité, ses encouragements et les précieux conseils prodigués tout le long de ce travail.

Je remercie le directeur du groupe de codage, le professeur Dr. Roch Lefebvre, pour son aide financière et ses suggestions remarquables qui ont été une source indispensable à la réalisation de ce projet de recherche.

J'exprime ma profonde reconnaissance à mes parents qui ont souffert pendant mon absence qui permettait cependant de continuer mes études.

Je tiens aussi à remercier mes frères Mohamed, Ali et Hussein pour leurs encouragements et leur soutien financier tout au long de mes études. Merci beaucoup!

Finalement, un grand merci à ma sœur Nohad, mon frère Dr. Walid Baalbaki et tous les ami(e)s qui m'ont encouragé à franchir cette étape.

Table des matières

DÉDICACE	
SOMMAIRE	
REMERCIEMENTS	
TABLE DES MATIÈRES	I
LISTE DES TABLEAUX	II
LISTE DES FIGURES	III
CHAPITRE 1	1
INTRODUCTION	1
CHAPITRE 2	6
ÉTAT DE L'ART	6
CHAPITRE 3	10
RÉDUCTEUR DE BRUIT POUR LA BANDE TÉLÉPHONIQUE	10
3.1 Introduction	10
3.2 Fenêtrage et DFT	12
3.2.1 Fenêtre de Hanning	12
3.2.2 La Transformée de Fourier Discrète	17
3.3 Estimation de l'énergie par bande	20
3.4 Détecteur Parole / Bruit, (DAV)	22
3.5 Estimation du RSB par bande	24
3.6 Modification du RSB par canal	25
3.7 Estimation du Bruit	26
3.8 Calcul du gain γ_c par bande	27
3.9 Filtrage dans le domaine fréquentiel	28
3.10 IDFT, fenêtrage, chevauchement et addition	29
3.10.1 Transformée de Fourier Discrète Inverse (IDFT)	29
3.10.2 Fenêtre sinusoïdale de reconstruction	31
CHAPITRE 4	32
RÉSULTATS	32
4.1 Introduction	32
4.2 Méthode du test	32
4.3 Résultats de tests d'écoute	33
CHAPITRE 5	42
RÉDUCTEUR DE BRUIT POUR LA LARGE BANDE	42
5.1 Introduction	42
5.2 Détails de l'algorithme	43
5.3 Fenêtrage et DFT	43
5.4 Estimation de l'énergie du canal	43
5.5 Détecteur Parole / Bruit, (DAV)	45
5.6 Estimation du RSB par canal	45

5.7	<i>Modification du RSB par canal</i>	45
5.8	<i>Calcul du gain γ_c par bande</i>	46
5.9	<i>Filtrage dans le domaine fréquentiel</i>	46
5.10	<i>Estimation du Bruit</i>	47
5.11	<i>IDFT, fenêtrage , chevauchement et addition (overlap Add)</i>	47
5.12	<i>Résultats</i>	47
5.12.1	<i>Statistiques</i>	56
CONCLUSION		58
BIBLIOGRAPHIE		60
ANNEXE A		63
LISTE DES SYMBOLES ET ABRÉVIATIONS		63

Liste des tableaux

Tableau 3-1	Fenêtrage utilisé par Ramabadran et al, et celui de l'algorithme proposé.	19
Tableau 3-2	Bandes critiques utilisée pour la voix à bande étroite.	21
Tableau 4-1	Pourcentage de votes pour des signaux de parole avec une voix d'homme à deux niveaux de bruit. Le nombre d'auditeurs est de treize.	33
Tableau 4-2	Pourcentage de votes pour des signaux de parole avec une voix de femme à deux niveaux de bruit. Le nombre d'auditeurs est de treize.	34
Tableau 4-3	Moyenne générale des pourcentages de votes pour les signaux de parole avec les voix d'hommes et de femmes à deux niveaux de bruit.	35
Tableau 5-1	Bandes critiques pour la bande élargie.	44
Tableau 5-2	Notations de l'écoute.	56

Liste des figures

Figure 3.1	Schéma-bloc d'un réducteur de bruit basé sur la soustraction spectrale et le RSB.	10
Figure 3.2	Chevauchement à 50% de deux fenêtres de Hanning.	13
Figure 3.3	Chevauchement à 50% de deux fenêtres sinusoïdales.	13
Figure 3.4	Deux fenêtres « cosinus surélevé », chacune de longueur 104.	15
Figure 3.5	Spectres des filtres passe-bandes modulés par les 3 fenêtres mentionnées.	16
Figure 3.6	Spectres de la figure 3.5 entre 0 et 1000 Hz.	17
Figure 3.7	Fenêtre sinusoïdale de longueur 256.	18
Figure 3.8	Fenêtre cosinus surélevé pour la DFT, la longueur de cette fenêtre est 104 plus 24 zéros.	18
Figure 3.9	Méthode d'un banc de filtres. H5 et H3 sont des filtres d'ordre respectif 5 et 3 [18].	23
Figure 3.10	Deux spectres de bruit avant et après traitement.	26
Figure 4.1	Préférence subjective des réducteurs de bruit pour un RSB de 10 dB.	35
Figure 4.2	Préférence subjective des réducteurs de bruit pour un RSB de 5 dB.	36
Figure 4.3	Réducteur de bruit. (a) Signal bruité avec un RSB de 10 dB. (b) Signal traité.	36
Figure 4.4	Parties voisées des signaux de parole de la figure 4.3.	37
Figure 4.5	Spectres d'amplitude des deux signaux de la figure 4.4.	37
Figure 4.6	Réducteur de bruit. (a) Trames du bruit seulement avec un RSB de 5 dB. (b) Signal traité.	38
Figure 4.7	Spectres d'amplitude des deux signaux présenté à la figure 4.6.	38
Figure 4.8	(a) Signal de parole propre original. (b) Signal traité avec l'algorithme proposé.	39
Figure 4.9	(a) Partie de signal de la figure 4.8 (a). (b) Partie de signal de la figure 4.8 (b).	40
Figure 4.10	Deux spectres confondus des deux signaux de la figure 4.8.	40
Figure 5.1	Réduction de bruit. (a) Signal de parole bruité avec un RSB de 5 dB. (b) Signal débruité par l'algorithme proposé.	48
Figure 5.2	(a) Signal de parole voisé bruité avec un RSB de 5 dB. (b) Signal traité.	48
Figure 5.3	Spectres d'amplitude des deux signaux voisés de la figure 5.2.	49

Figure 5.4	Réduction de bruit. (a) Bruit de bafouille avec un RSB de 5 dB. (b) Bruit traité par l'algorithme proposé pour la bande élargie.	50
Figure 5.5	Spectres d'amplitude des deux signaux de la figure 5.4.	50
Figure 5.6	Réduction de bruit. (a) Bruit des voitures et de la rue avec un RSB de 10 dB. (b) Bruit traité.	51
Figure 5.7	Spectres d'amplitude des deux signaux de la figure 5.6.	52
Figure 5.8	(a) Signal de parole propre. (b) Signal traité.	52
Figure 5.9	Spectres d'amplitude des deux signaux de la figure 5.8.	53
Figure 5.10	(a) Signal propre de parole voisée. (b) Signal traité.	53
Figure 5.11	Deux spectres d'amplitude confondus pour les signaux de la figure 5.10.	54
Figure 5.12	(a) Partie non voisée du signal de parole de la figure 5.9 (a). (b) Signal traité.	54
Figure 5.13	Spectres d'amplitude des deux signaux de la figure 5.12.	55

Chapitre 1

Introduction

Le but de ce mémoire est de présenter un algorithme qui réduit le bruit dans un signal de parole bruité par des bruits additifs et non corrélés au signal de parole propre. Un environnement bruité nuit à la clarté d'un signal de parole propre; ce qui rend ce signal ennuyant à l'oreille humaine. Dans la situation où le niveau de bruit est élevé, le signal de parole propre devient incompréhensible. La question qui se pose à priori est de savoir l'origine de ce bruit dérangeant. Plusieurs éléments peuvent intervenir à créer du bruit qui s'additionne au signal de parole propre. Par exemple :

- 1- Une communication établie avec un téléphone cellulaire dans une voiture où l'air, le moteur et les voitures au voisinage font du bruit qui vient s'ajouter au signal de parole original, ce qui nuit à sa clarté.
- 2- Le signal propre correspondant à une communication entamée à partir d'une cabine téléphonique dont l'environnement est très bruité avec le passage de voitures, de camions et d'avions, est affecté tout en provoquant un malaise auditif aux deux bouts de la ligne téléphonique.
- 3- Le bruit qui résulte de l'activation de l'ordinateur et de ses composantes (clavier, imprimante, scanner,...), du fonctionnement d'une lampe halogène, affecte la clarté d'une conversation téléphonique au bureau.

- 4- Dans les applications militaires, l'environnement est particulièrement bruité (chars, hélicoptères, avions de chasse, etc.) ce qui perturbe de façon importante la qualité des communications vocales. Le niveau de bruit peut même atteindre ou dépasser le niveau du signal de parole (RSB nul ou négatif).

Ce problème dans les systèmes de communications a créé un besoin réel pour améliorer les performances de ceux-ci en créant des algorithmes qui seront capables de fonctionner même dans les cas où le RSB est faible. Pour se faire, divers algorithmes et méthodes ont été proposés.

Les algorithmes de réduction du bruit tentent d'améliorer deux caractéristiques propres au signal de parole bruité : la première, qualitative, est une mesure subjective qui montre jusqu'à quel niveau le son du signal débruité plaît aux auditeurs. L'amélioration de la qualité sonore n'est pas seulement du plaisir esthétique, mais de la facilité de compréhension afin de minimiser la fatigue pour l'oreille humaine. La deuxième caractéristique, intelligibilité, est une mesure objective qui montre la quantité de l'information que nous sommes capables d'extraire du signal de parole bruité sans se soucier de sa qualité.

Les algorithmes qui se basent sur la soustraction spectrale ou le filtrage spectral constituent l'approche traditionnelle pour diminuer le bruit additif. Dans cette technique, on estime le signal de parole débruité dans le domaine fréquentiel. Cette estimation se base sur la densité de puissance spectrale du bruit déduit du spectre de signal bruité. La densité de puissance spectrale du bruit est mesurée pendant les périodes de silence, en absence du signal de parole.

La soustraction spectrale est une technique de base qui a été utilisée dans plusieurs algorithmes qui réduisent le bruit additif [1]. La popularité de ces types d'algorithmes est basée sur leur

flexibilité et leur simplicité d'implémentation. Cependant, l'inconvénient de la soustraction spectrale est l'introduction du bruit appelé « bruit musical résiduel » avec une structure non naturelle. Ce bruit perceptible et ennuyant est composé de tonalités à des fréquences arbitraires avec une variance élevée.

Pour résoudre ce problème de bruit musical, plusieurs solutions ont été proposées [2, 16]. Voici quelques unes: Amplitude moyenne [1], soustraction spectrale non linéaire [3], soustraction du bruit en introduisant un planché de bruit dans le spectre [4], estimation optimale de MMSE par l'amplitude spectrale à court terme [6].

Cependant, dans le cas où le RSB est faible, des tests ont démontré que ces algorithmes nécessitent des améliorations notables. Ainsi, diverses méthodes ont été proposées par les chercheurs telles que le rehaussement de la parole en se basant sur les propriétés de masquage du système d'écoute humaine [5] ou en utilisant le critère psycho-acoustique [7], l'extraction des paramètres de la parole dans un environnement bruité en utilisant les propriétés de masquage [8] et les propriétés de sous-espace d'un signal de parole bruité [17].

Ainsi, le bruit musical résiduel causé par la soustraction spectrale et le faible niveau du RSB constituent deux facteurs qui exigent un algorithme efficace qui sera capable de les résoudre sans pour autant créer des distorsions dans le signal de parole original. La plupart des algorithmes proposés traitent des signaux pour la bande téléphonique de 4 kHz. Avec l'avènement de nouvelles technologies, de nombreux services dans le domaine de la téléphonie cellulaire, telle que la téléconférence, sont offerts aux utilisateurs. Une qualité de voix plus élevée étant requise, nous avons décidé de réaliser un réducteur du bruit pour la bande élargie.

Le premier objectif fixé dans ce travail est d'optimiser l'algorithme de Ramabadran et al (RAML, Ramabadran, Ashley et McLaughlin), pour le rendre efficace. Pour ce faire, nous avons pris l'idée générale de Ramabadran et al, et nous avons écrit un nouveau code qui contient nos modifications.

Le deuxième but de ce travail consiste à étendre cet algorithme pour les signaux à bande élargie. Pour cela, nous proposons deux algorithmes qui se basent sur la soustraction spectrale et le RSB. Un algorithme traite les signaux pour la bande étroite de 4 kHz et l'autre traite les signaux pour la bande élargie à 8 kHz.

Le mémoire est organisé comme suit :

Le chapitre 2 du présent mémoire, que nous avons appelé « état de l'art », présente l'historique de la réduction du bruit (qui a commencé en 1970 avec M. Boll [1]) et les recherches actuelles dans ce domaine.

Dans le chapitre 3, nous présentons l'algorithme pour les signaux dans la bande téléphonique de 4 kHz avec toutes les étapes et les essais qui ont conduit à un réducteur du bruit efficace et qui respecte les conditions posées au départ.

Ensuite, le quatrième chapitre présente les tests et les résultats qui ont été faits sur cet algorithme et les comparaisons établies avec l'algorithme initial de Ramabadran et al.

Au chapitre 5, nous présenterons une extension de l'algorithme pour les signaux à bande élargie (i.e. avec une fréquence d'échantillonnage de 16 kHz). Comparativement aux signaux à bande étroite (échantillonnés à 8 kHz), les signaux à bande élargie ont une dynamique spectrale plus

importante et leur décomposition en bandes critiques (une des bases de l'algorithme proposé) produit un nombre plus important de bandes. De plus, le détecteur d'activité vocale (DAV) doit être optimisé pour ces signaux. Le chapitre 5 présentera donc des modifications de l'algorithme pour l'appliquer aux signaux à bande élargie, et donnera les résultats de tests subjectifs qui ont été réalisés.

Chapitre 2

État de l'art

Depuis trois décennies, des recherches ont été effectuées pour extraire le bruit additif du signal de parole originale (signal bruité). Les recherches dans ce domaine s'accroissent depuis l'arrivée de la téléphonie cellulaire, de la téléconférence et de la reconnaissance de la parole, car dans ces situations, le signal de parole est confronté à un niveau de bruit élevé qui peut nuire à la qualité de service et à la performance des systèmes.

La réduction du bruit a été popularisée en 1979 par la technique de soustraction spectrale présentée par M. Boll et d'autres [1].

Un autre modèle de réducteur de bruit a été développé par J. S. Lim et A. V. Oppenheim dans un article publié en 1978 [14]. Le signal de parole y est modélisé comme un processus autorégressif variant dans le temps qui sera estimé du signal de parole bruité. En pratique, cette approche n'a pas bien fonctionné, parce que les échantillons du signal bruité seuls ne contiennent pas assez d'informations pour estimer exactement les paramètres du modèle et le signal lui-même.

Un an plus tard, en 1979, Lim et Oppenheim ont publié un article [2] dans lequel ils ont examiné différents systèmes de rehaussement de la parole connus jusqu'à date. Les auteurs ont conclu que

les systèmes basés sur la soustraction spectrale sont les plus efficaces pour améliorer le signal de parole bruité par du bruit additif et non-corrélié au signal de parole propre.

Par la suite, les chercheurs ont développé plusieurs versions qui se basent sur la soustraction spectrale. Toutefois, le bruit restant devient « musical » et nuisible subjectivement.

Pour cela, en 1980 R. J. McAulay et M. L. Malpass [13] ont tenté d'améliorer la performance de la soustraction spectrale. Ils se sont basés sur une idée introduite par Drucker [12] en 1968. Drucker a utilisé différents estimateurs pour cinq classes de signal de parole. Dans son approche, il explique que le signal de parole propre peut se présenter en divers états. Il a utilisé de multiples filtres conçus spécialement pour chaque état. McAulay et Malpass ont amélioré les performances de l'approche de la soustraction spectrale en appliquant le même concept avec un chemin différent. Ce concept se repose sur deux états de parole, silence et non silence.

Y. Ephraïm combine les idées fondamentales de Drucker, McAulay, Malpass, Lim et Oppenheim dans un article publié en 1992, [15].

En 1995, Y. Ephraïm et H. L. Van Trees proposent une méthode autre que celle de la soustraction spectrale pour éliminer ce bruit musical [17]. Ils décomposent l'espace vectoriel du signal bruité à un sous-espace du signal plus du bruit et à un sous-espace de bruit. La réduction de bruit est améliorée en soustrayant le sous-espace du bruit et en estimant le signal débruité de ce qui reste du sous-espace de signal. Malgré un concept avancé, la performance de cette méthode est comparable à celle de la soustraction spectrale.

En 1995, N. Virag a présenté un article qui se base sur la soustraction spectrale et les propriétés de masquage de l'oreille humaine [5]. L'idée consiste à soustraire moins de bruit dans les endroits où le seuil de masquage est élevé. Bien entendu, dans le cas où le seuil de masquage est élevé, le bruit n'est pas audible. Il est à noter que les paramètres de cet algorithme sont adaptatifs.

Une approche différente a été proposée dans un article présenté en 1997 par Ramabadran, Ashley et McLaughlin [11]. Cette méthode repose sur la soustraction spectrale par bande et le RSB. Le spectre d'amplitude est décomposé en plusieurs bandes critiques pour lesquelles le RSB est calculé individuellement. Globalement, nous pouvons dire que la méthode de Ramabadran et al a les mêmes principes que ceux de Virag, toutefois Virag se base sur le seuil de masque pour décider du niveau de bruit à soustraire, alors que Ramabadran et al se fondent sur le niveau du RSB par bande spectrale. Les éléments fondamentaux dans de la méthode de Ramabadran et al sont :

- Conversion du signal de parole dans le domaine de la fréquence;
- Décomposition du spectre de signal en bandes critiques;
- Calcul du RSB dans chaque bande critique;
- Calcul d'un gain compris entre 0 et 1 qui multiplie le spectre bruité.
- La valeur du gain dépend de la valeur du RSB et de la décision d'un détecteur de l'activité vocale, DAV.

Les majeurs défauts de l'algorithme de Ramabadran et al sont la modification éventuelle du signal de parole propre et l'efficacité limitée pour certains types de bruit comme celui de bafouille.

Dans le but d'améliorer subjectivement l'algorithme de Ramabaran et al, le chapitre 3 de ce présent travail traitera plusieurs changements apportés :

Le DAV, le calcul du gain par bande, la méthode de modification du RSB et aussi le fenêtrage.

Ensuite, dans le chapitre 5, les mêmes techniques seront généralisées pour les signaux à bande élargie.

Chapitre 3

Réducteur de bruit pour la bande téléphonique

3.1 Introduction

Le but principal est de présenter un algorithme qui réduit le bruit dans la bande téléphonique de 4 kHz. Le schéma-bloc de la figure 3.1 illustre les parties requises pour un algorithme basé sur la soustraction spectrale et sur le rapport signal sur bruit.

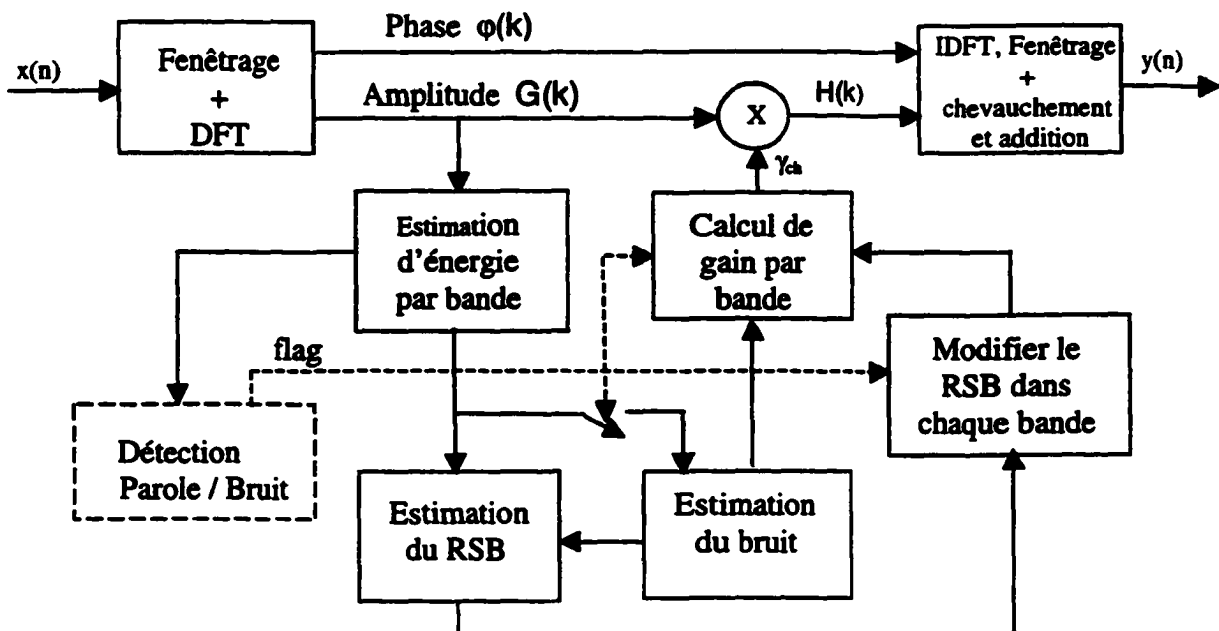


Figure 3.1 Schéma-bloc d'un réducteur de bruit basé sur la soustraction spectrale et le RSB.

À la sortie de ce schéma-bloc, nous trouvons le signal $y(n)$ dont le niveau du bruit est réduit par rapport au signal d'entrée, $x(n)$. Le signal d'entrée, $x(n)$, est échantillonné à 8000 Hz. Ensuite, ce signal est fenêtré par une racine carrée de la fenêtre de Hanning puis transformé dans le domaine fréquentiel par le biais d'une transformée de Fourier discrète (DFT). À ce stade, le signal est représenté en phases et amplitudes. Les phases sont gardées intactes, les modifications s'effectuent seulement sur la puissance spectrale. Ensuite, le spectre d'amplitude est décomposé en plusieurs bandes critiques ou plusieurs canaux. À partir de ces bandes critiques, nous calculons l'énergie lissée par bande. Ainsi, nous calculons l'énergie lissée du bruit dans le cas où le DAV (Détection de l'Activité Vocale) détecte une trame du bruit. Nous calculons alors le RSB par bande. Le RSB calculé sera modifié suivant son niveau et son emplacement dans les bandes critiques. Par exemple, si le RSB est faible et que nous sommes en présence du bruit, le RSB sera remis à 1 dB dans toutes les bandes critiques. Le but de modifier le RSB sert à soustraire plus de bruit dans le cas où le RSB faible. La dernière étape du calcul consiste à trouver le meilleur gain multipliant chacune des bandes critiques du spectre bruité. Finalement, le spectre du signal bruité sera modifié et converti dans le domaine du temps en utilisant la Transformée de Fourier Discrète Inverse (IDFT). Le fenêtrage à la synthèse ou à la reconstruction par la racine carrée de la fenêtre de Hanning est essentiel pour que la multiplication et l'addition de deux fenêtres chevauchées (analyse et synthèse) donnent 1.

Il faut mentionner que pour réaliser ce projet, plusieurs tests d'écoute ont été faits avec différents niveaux de bruit et différents genres de bruit pour s'assurer que ce réducteur est efficace dans toutes les situations où le signal de parole original est détérioré par ce genre de bruit. Ensuite, des changements de paramètres ont été nécessaires pour optimiser cet algorithme.

Pour comprendre plus en détail le principe de fonctionnement de ce réducteur, nous présentons, dans les sections à venir, les détails de chaque bloc du schéma de la figure 3.1.

3.2 Fenêtrage et DFT

Le signal de parole $x(n)$ qui se trouve à l'entrée du schéma-bloc est échantillonné à 8 kHz. En premier lieu, nous prenons des trames de 128 échantillons, 16 ms. La première étape de calcul consiste à fenêtrer $x(n)$ par la racine carrée de la fenêtre de Hanning; ce qui est équivalent à une fenêtre sinusoïdale. Ensuite, nous transformons ce signal fenêtré dans le domaine fréquentiel, par une DFT de 256 points.

3.2.1 Fenêtre de Hanning

Une des méthodes possibles pour limiter autant que possible l'effet de fuite, lors de l'application de la DFT à des signaux périodiques, consiste à utiliser une « fenêtre ». Le choix de la fenêtre sinusoïdale est basé sur plusieurs essais par rapport à d'autres fenêtres comme celles de Hamming, rectangulaire, de Hanning et celle proposée par Ramabadran et al. La longueur de la fenêtre utilisée est deux fois plus grande que la longueur de trame (256 échantillons). Donc, ceci présente un chevauchement de 50%.

La figure 3.2 présente deux fenêtres de Hanning chevauchées à 50%.

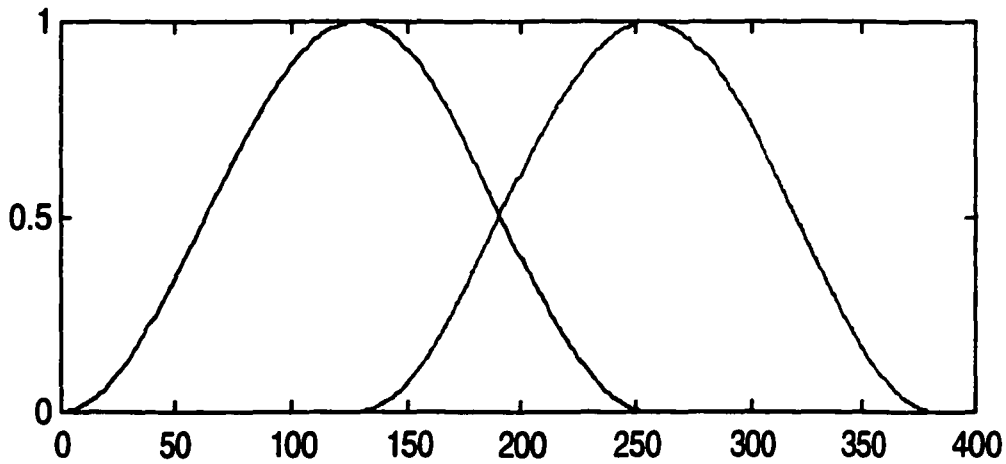


Figure 3.2 Chevauchement à 50% de deux fenêtres de Hanning.

Si nous utilisons la fenêtre de Hanning comme fenêtre d'analyse, la fenêtre de reconstruction sera une fenêtre rectangulaire.

Suite à des tests comparant la fenêtre de Hanning, la fenêtre sinusoïdale et celle proposée par Ramabadran et al, nous avons constaté que la fenêtre sinusoïdale est la plus appropriée au niveau de l'audition humaine. La figure 3.3 présente deux fenêtres sinusoïdales de longueur 256 avec un chevauchement de 50%.

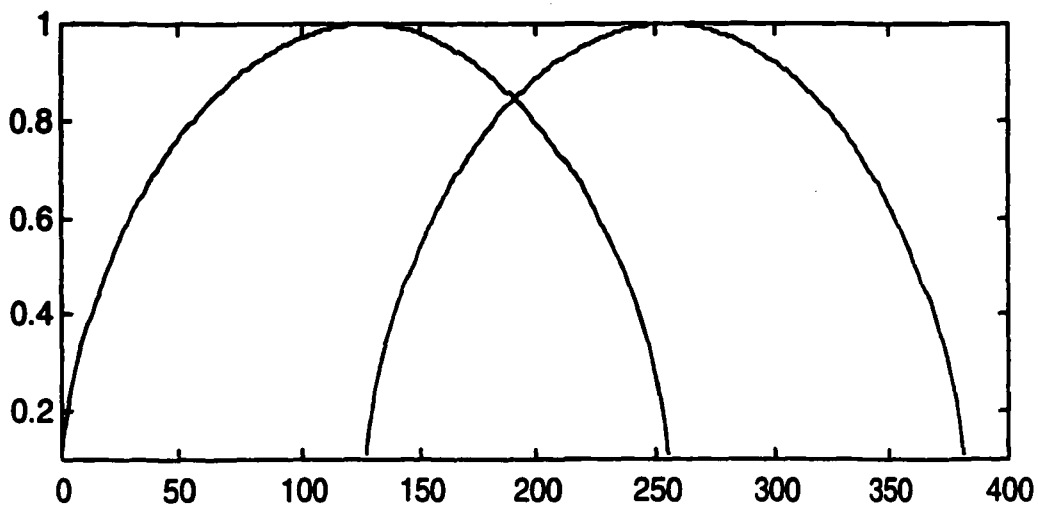


Figure 3.3 Chevauchement à 50% de deux fenêtres sinusoïdales.

La racine carrée de la fenêtre de Hanning est donnée par:

$$w(i) = \sqrt{0.5 - 0.5 \cos\left(\frac{2\pi}{L}i\right)} \quad \text{pour } i = 0, 1, \dots, L-1 \quad (3-1)$$

où L est la longueur de la fenêtre, $L = 256$.

L'équation 3-1 de la racine carrée de la fenêtre de Hanning peut être réécrite sous forme d'équation d'une fenêtre sinusoïdale comme suit :

$$w(i) = \sqrt{0.5 - 0.5 \cos\left(\frac{2\pi}{L}i\right)} = \sqrt{0.5 - 0.5\left(1 - 2 \sin^2\left(\frac{\pi}{L}i\right)\right)} = \sin\left(\frac{\pi}{L}i\right) \quad (3-2)$$

où $i = 0, 1, \dots, L-1$

Le signal d'entrée fenêtré, $x_f(i)$, est donné par :

$$x_f(i) = x_{\text{buf}}(i) * w(i) \quad 0 \leq i < L \quad (3-3)$$

$x_{\text{buf}}(n)$ contient les 128 échantillons de la trame présente et les 128 échantillons de la trame précédente.

L'utilisation de la fenêtre sinusoïdale pour fenêtrer $x_{\text{buf}}(n)$ exige que nous fenêtrons ce signal après traitement; ce que nous appelons la fenêtre de reconstruction. Pour que la fenêtre sinusoïdale soit respectée, il faut la multiplier une deuxième fois par le signal fenêtré à l'entrée et l'additionner (overlap Add). Le but de ce multiplication et de cette addition est de ramener les deux fenêtres chevauchées (analyse et synthèse) à 1. Le fait d'utiliser une fenêtre sinusoïdale à l'analyse et à la synthèse, nous donne des résultats meilleurs que ceux de Ramabadran et al pour des raisons qui vont être élucidées par la suite. La figure 3.4 montre le principe de fenêtrage qui a été utilisé par Ramabadran et al [11] :

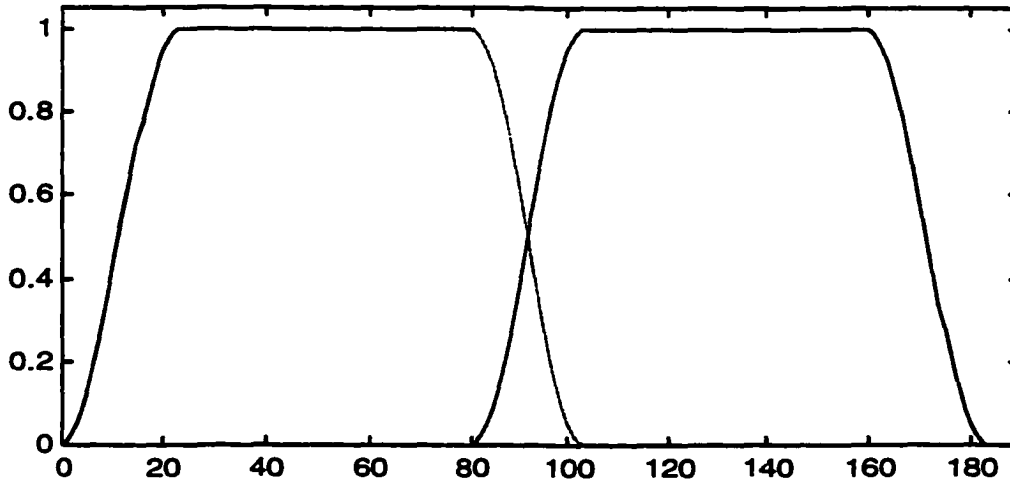


Figure 3.4 Deux fenêtres « cosinus surélevé », chacune de longueur 104.

L'équation de la fenêtre représentée à la figure 3.4 est la suivante :

$$w_M(n) = \begin{cases} \sin^2(\pi(n+0.5)/2D) & 0 \leq n < D \\ 1 & D \leq n < L_M \\ \sin^2(\pi(n-L_M+D+0.5)/2D) & L_M \leq n < D+L_M \\ 0 & D+L_M \leq n < M \end{cases} \quad (3-4)$$

Avec $D = 24$, $L_M = 80$ et $M = 128$.

Cette fenêtre est utilisée avec un chevauchement de 24 échantillons.

La figure 3.5 illustre les effets de bord causés par chacune des fenêtres mentionnées ci-dessus (la fenêtre sinusoïdale, la fenêtre de Hanning et la fenêtre cosinus surélevé).

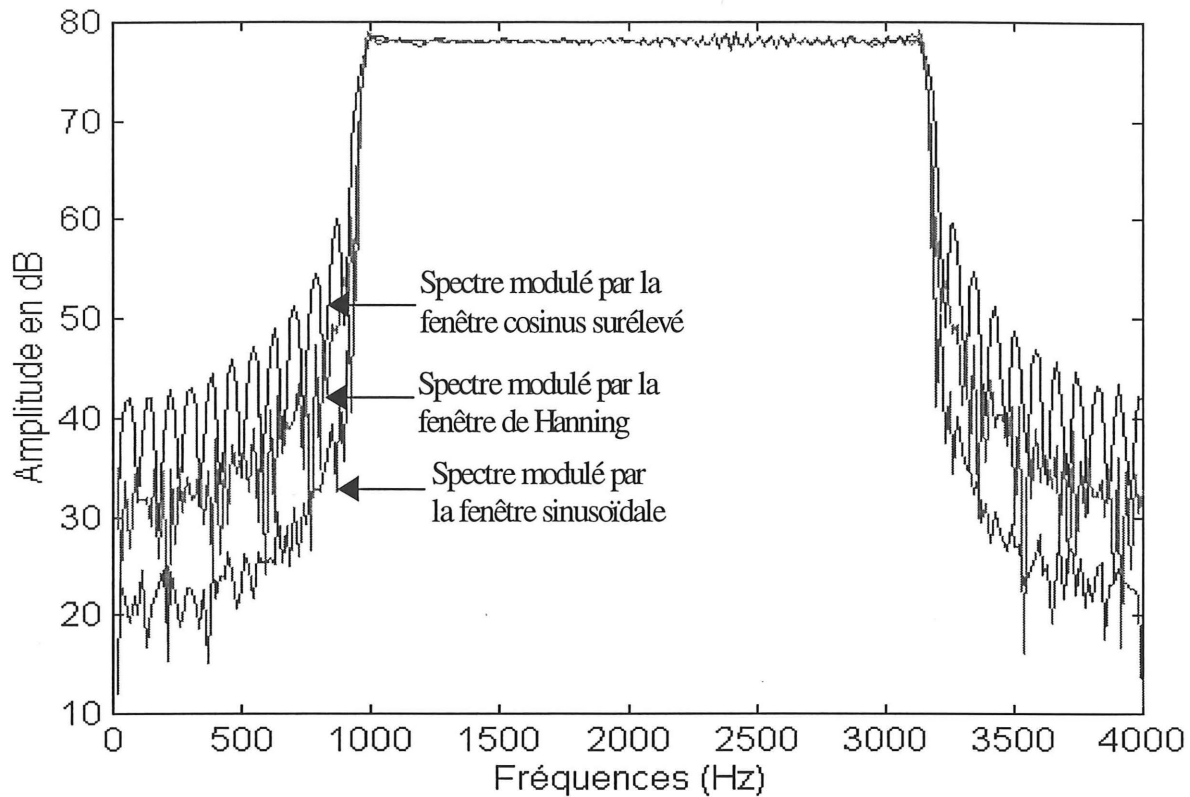


Figure 3.5 Spectres des filtres passe-bandes modulés par les 3 fenêtres mentionnées.

Dans cette figure, l'utilisation de la fenêtre sinusoïdale à l'analyse et à la synthèse donne une meilleure atténuation (atténuation = $78 - 20 = 58$ dB) que ceux des autres fenêtres (48 dB avec la fenêtre de Hanning et 38 dB avec le cosinus surélevé).

La figure 3.6 présente les trois spectres de la figure 3.5 entre 0 et 1000 Hz. Cette figure montre de façon plus claire les effets de bord causés par chacune des fenêtres.

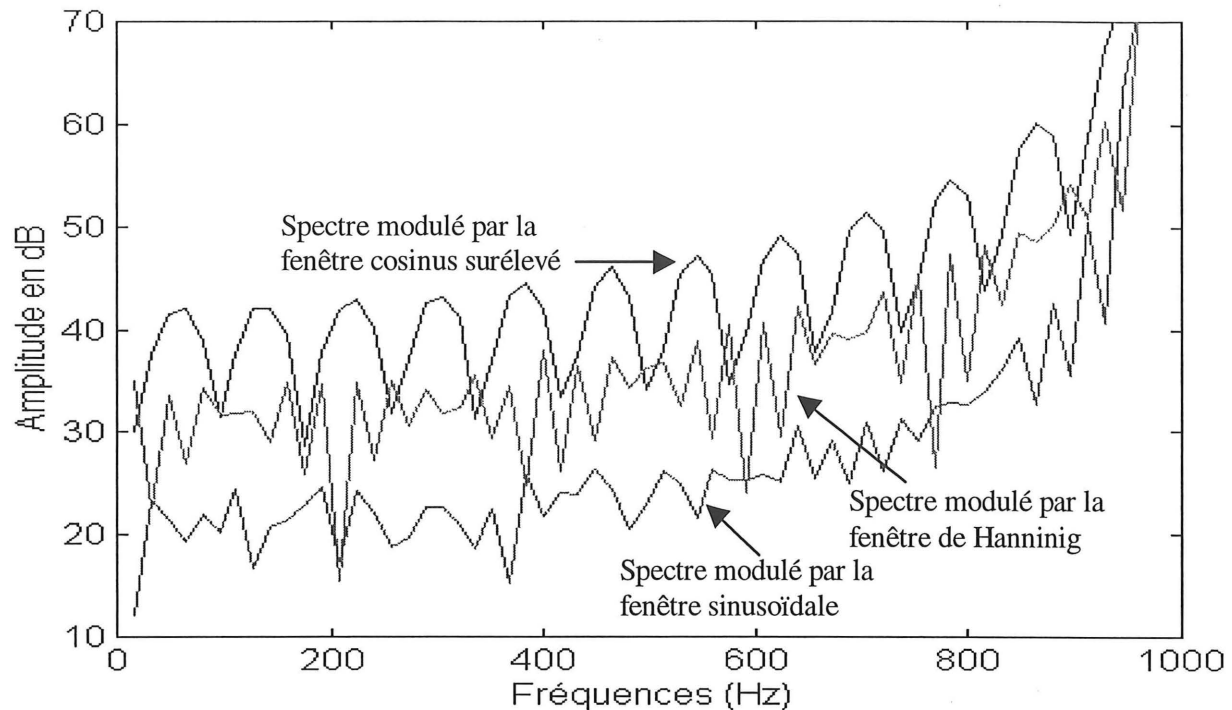


Figure 3.6 Spectres de la figure 3.5 entre 0 et 1000 Hz.

Nous voyons de façon très claire que la fenêtre choisi par Ramabadran et al cause des effets de bord les plus élevés sur le spectre d'amplitude et avec plus d'ondulations. Il est à noter que des tests d'écoute ont montré que le choix de la fenêtre sinusoïdale est le meilleur. La différence dans le spectre apportée par chacune des fenêtres est presque de 10 dB.

3.2.2 La Transformée de Fourier Discrète

La transformation du signal $x(n)$ dans le domaine fréquentiel est nécessaire pour que nous puissions travailler dans le domaine spectral. Ceci est réalisé par l'intermédiaire d'une DFT de L points. La longueur de la fenêtre sinusoïdale est de 256. Cette fenêtre est présentée à la figure 3.7.

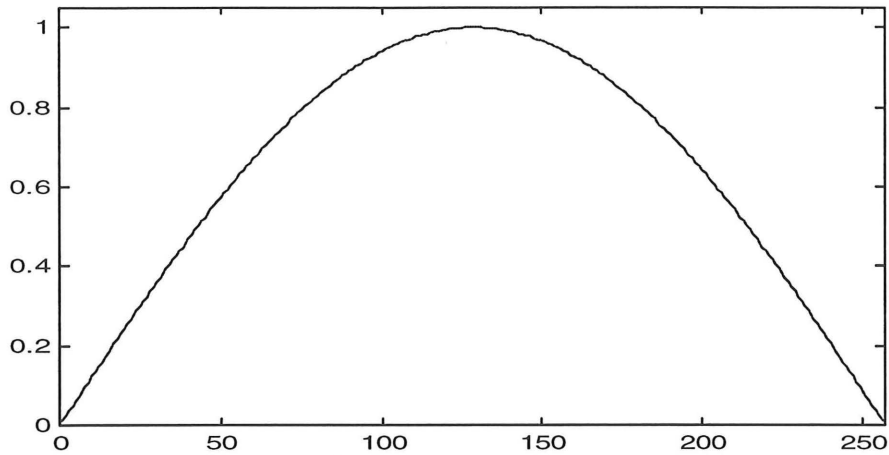


Figure 3.7 Fenêtre sinusoïdale de longueur 256.

Dans le cas de Ramabadran et al, le fenêtrage utilisé pour faire la DFT rapide est présenté à la figure 3.8.

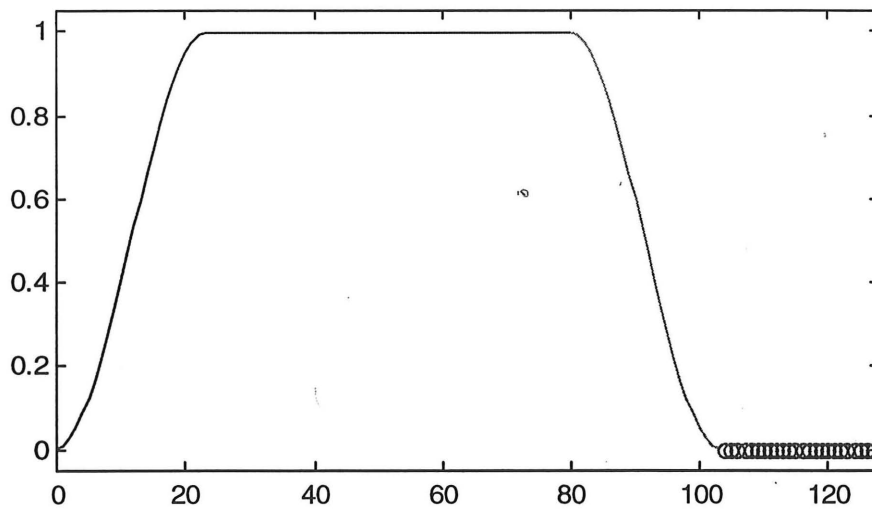


Figure 3.8 Fenêtre cosinus surélevé pour la DFT, la longueur de cette fenêtre est 104 plus 24 zéros.

Pour faire une DFT rapide, le nombre de points doit être une puissance de 2. Donc, puisque la longueur de la fenêtre est de 104, l'insertion de 24 zéros est nécessaire. Ceci augmente le nombre de points, mais pas la résolution.

Rendu à cette étape, nous pouvons discuter d'autres avantages de l'utilisation de la fenêtre sinusoïdale comparativement à celle cosinus surélevé par rapport à la DFT. Le tableau 3-1 présente une comparaison entre les deux fenêtres.

Tableau 3-1 Fenêtrage utilisé par Ramabadran et al, et celui de l'algorithme proposé.

	Fenêtre sinusoïdale	Fenêtre cosinus surélevé
Longueur de la fenêtre	256	104 + 24 zéros = 128
Fenêtrage	À l'analyse et au synthèse	À l'analyse
Forme	Avec atténuation au milieu	Plat au milieu

En premier lieu, nous discutons de la longueur de la fenêtre : la longueur de la fenêtre sinusoïdale est de 256 échantillons par rapport à 104 échantillons et 24 zéros qui ont été ajoutés pour faire une DFT rapide dans le cas de la fenêtre cosinus surélevé. Alors, dans notre cas il y a plus de résolution (2 fois plus).

Ensuite, nous parlons de l'effet du double fenêtrage : le fenêtrage à l'analyse et au synthèse évite le problème du bruit indésirable causé par les effets de bord du fenêtrage; ce qui augmente la qualité sonore.

Finalement, nous analysons la forme de la fenêtre: la forme plate au milieu de la fenêtre cosinus surélevé laisse le signal d'entrée sans aucun changement; ce qui diminue l'effet du fenêtrage. Outre les raisons précédentes, le choix de la fenêtre sinusoïdale a été basé sur un petit test d'écoute.

La définition d'une DFT de L points est la suivante :

$$X(k) = \frac{1}{\sqrt{L}} \sum_{n=0}^{L-1} x_f(n) e^{(-j\frac{2\pi nk}{L})} \quad (3-5)$$

$X(k)$ représente le signal $x_f(n)$ dans le domaine de fréquence .

La présence du $\frac{1}{\sqrt{L}}$ a pour objectif d'égaliser l'énergie dans le domaine du temps et dans le domaine de fréquence.

L'équation 3-5 peut se réécrire comme suit :

$$X(k) = \frac{1}{\sqrt{L}} \sum_{n=0}^{L-1} x_f(n) \left(\cos\left(\frac{2n\pi k}{L}\right) - j \sin\left(\frac{2n\pi k}{L}\right) \right) \quad (3-6)$$

où $k = 0, 1, 2, \dots, L-2, L-1$ et $L = 256$.

À partir de l'équation 3-6, nous pouvons calculer les amplitudes et les phases de $X(k)$. Nous gardons les phases jusqu'à la fin de nos calculs sans les changer. Par contre, tout le traitement va se concentrer sur le spectre d'amplitude.

3.3 Estimation de l'énergie par bande

Avant d'estimer l'énergie par canal ou l'énergie par bande pour chaque trame m , nous décomposons le spectre d'amplitude en 18 bandes critiques, $B(i)$, avec $0 \leq i < 18$. Cette façon de faire est basée sur l'article de J. D. Johnston, [10]. Le principe consiste à rassembler plusieurs raies dans une même bande en faisant leur moyenne. Donc, toutes les raies ont la même valeur dans une bande. Ceci aide à convertir les raies spectrales en bandes critiques qui ne varient pas de façon instantanée comme dans le cas des raies spectrales individuelles. De plus, pour ne pas déformer le spectre d'amplitude, le nombre de raies rassemblées dans les bandes à basses fréquences est moins élevé que celui dans les hautes fréquences. Car, la déformation du spectre dans les basses fréquences est plus audible que dans les hautes fréquences. Le tableau 3-2 décrit le nombre de raies existantes dans chaque bande critique $B(i)$.

Tableau 3-2 Bandes critiques utilisées pour la voix à bande étroite.

Bande critique i ^{ème} nombre,	Les raies de la DFT		Fréquences réelles [Hz]
	Intervalles	Nombre de raies	
1	1-3	3	0-94
2	4-6	3	94-187
3	7-10	4	187-312
4	11-13	3	312-406
5	14-16	3	406-500
6	17-20	4	500-625
7	21-25	5	625-781
8	26-29	4	781-906
9	30-35	6	906-1094
10	36-41	6	1094-1281
11	42-47	6	1281-1469
12	48-55	8	1469-1719
13	56-64	9	1719-2000
14	65-74	10	2000-2312
15	75-86	12	2312-2687
16	87-100	14	2687-3125
17	101-118	18	3125-3687
18	119-128	10	3687-4000

Le calcul de l'énergie dans chaque bande critique B(i) dans la première trame se fait suivant l'équation suivante :

$$E_c(m, i) = \sum_{i=0}^{N_b-1} \sum_{k=0}^{f_n[i]} \frac{(G[f_L[i] + k])^2}{f_n[i]} \quad \text{avec } m=1 \text{ et } 0 \leq N < N_b \quad (3-7)$$

où N_b est le nombre de bandes B(i), $N_b = 18$.

$G(k)$ est le spectre d'amplitude.

$f_L[i]$ est la partie inférieure de l'intervalle :

$f_L[i] = \{0, 3, 6, 10, 13, 16, 20, 25, 29, 35, 41, 47, 55, 64, 74, 86, 100, 118\}$.

$f_n[i]$ est le nombre de raies dans chaque bande :

$f_n[i] = \{3, 3, 4, 3, 3, 4, 5, 4, 6, 6, 6, 8, 9, 10, 12, 14, 18, 10\}$.

Pour $m > 1$, l'énergie de chaque bande critique est lissée par un facteur $\alpha_m = 0.45$ par un filtrage autorégressif.

L'équation 3-8 décrit la formulation de cette énergie :

$$E_c(m, i) = \max\{E_{\min}, \alpha_m E_c(m-1, i) + (1-\alpha_m) E_c(m, i)\} \quad 0 \leq N < N_b \text{ et } m \geq 2 \quad (3-8)$$

où $E_{\min} = 1/10 = 0.1$ est le minimum d'énergie allouée pour chaque canal. Car, la somme de l'énergie du bruit sur les 10 bandes critiques à l'équation 3-12 ne doit jamais être inférieur à 1.

La valeur de α_m est flexible. Le choix du 0.45 est approprié puisque le lissage dans ce cas-ci garde 0.45 de la trame passée et 0.55 de la trame actuelle. Nous pouvons choisir un autre lissage comme 0.4 du passé est 0.6 du présent. Ceci crée un lissage plus rapide.

Les calculs de l'énergie et les valeurs du E_{\min} et α_m sont inspirés de l'article [11] de Ramabadran et al.

3.4 Détecteur Parole / Bruit, (DAV)

La détection d'activité vocale (DAV) est largement utilisée dans les systèmes de communications. L'utilisation du DAV a pour but de signaler la présence d'une trame de bruit. Dans le cas d'absence de la parole, les systèmes de communications réduisent leur débit de transmission. Donc, la présence d'un DAV dans les systèmes de communications a pour objectif d'augmenter leur efficacité spectrale.

Le DAV est aussi nécessaire pour les algorithmes d'annulation du bruit et de l'écho acoustique qui ont besoins de calculer le model du bruit. Dans notre cas, l'objectif d'un DAV est de mettre à jour les caractéristiques du bruit pendant les périodes du silence. Dans la situation où le DAV

détecte une trame de bruit, nous calculons l'allure spectrale et l'énergie de ce bruit et nous les mettons à jour pour la prochaine trame, $m+1$.

Le défaut majeur d'un DAV est un mal fonctionnement éventuel.

Plusieurs méthodes ont été suggérées pour réaliser la fonctionnalité du DAV. La plupart de ces méthodes sont basées sur l'estimation du spectre du signal, [20] [21] [22]. Le spectre peut être estimé en utilisant la DFT, le banc de filtres, les coefficients de prédiction linéaire ou les LSF [22]. Dans notre cas, la base initiale de ce DAV se trouve dans l'article [18], qui représente un algorithme suggéré pour le système GSM AMR (Adaptive Multi-Rate). Dans cet article, les auteurs ont choisi de décomposer le signal en 9 bandes critiques en utilisant un banc de filtres. La décomposition du signal par filtrage [18], est présentée à la figure 3.9.

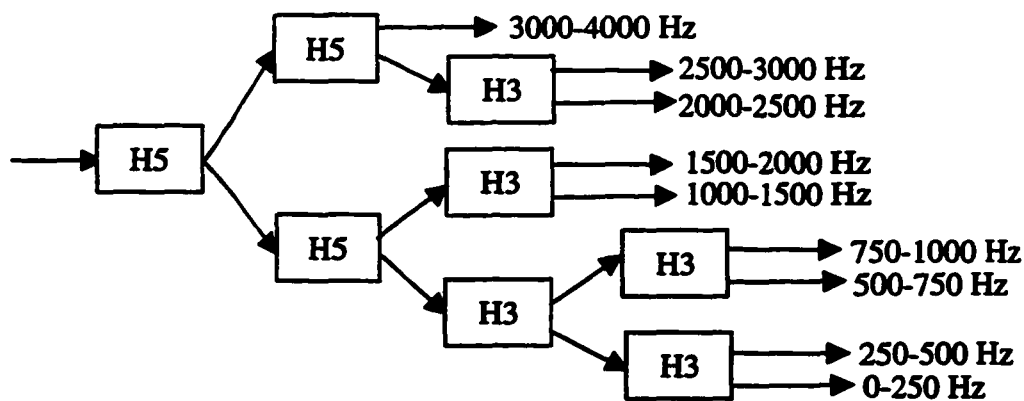


Figure 3.9 Méthode d'un banc de filtres. H5 et H3 sont des filtres d'ordre respectif 5 et 3 [18].

Les deux filtres, H5 et H3, sont des filtres qui divisent la bande spectrale d'un signal en deux parties ($0-\pi/2$ et $\pi/2-\pi$). Les deux filtres, H5 et H3, sont respectivement d'ordre 5 et 3. La structure du DAV est basée sur l'utilisation des propriétés spectrales du signal. Le bruit est

estimé séparément dans chaque sous-bande et la mise à jour de chaque composante estimée du bruit est contrôlée individuellement. En plus, le DAV utilise un détecteur de « pitch » en boucle ouverte pour détecter les tonalités de signalisations et d'autres signaux qui ont des composantes périodiques importantes.

Dans notre cas, puisque le spectre d'amplitude a été déjà calculé par la DFT, nous n'utilisons pas la méthode d'un banc de filtres. Par contre, au lieu de calculer 9 bandes critiques nous en calculons 18 selon le tableau 3-2.

Le mécanisme de décision, parole/bruit, est décrit dans la référence [18].

Le DAV a été inséré dans l'algorithme proposé sans modification majeure sauf dans le traitement de bandes critiques.

3.5 Estimation du RSB par bande

Le calcul du RSB est essentiel, puisqu'il nous aide à donner une pondération sur la quantité de bruit à soustraire dans chaque bande. Le calcul du RSB dans chaque bande se base sur l'énergie du signal et celle du bruit qui est estimé dans chaque trame pendant les périodes de silence.

Donc, le calcul du RSB en dB se fait comme suit :

$$RSB[i] = \max (0, \min(33, 10 \cdot \log_{10}(E_c[i]/E_b[i]))) \quad 0 \leq i < N_b \quad (3-9)$$

où E_b est l'énergie du bruit par bande.

La limitation du RSB entre 0 et 33 dB est essentielle afin de mieux contrôler les valeurs à l'intérieur de cet intervalle.

3.6 Modification du RSB par canal

Rendu à ce stade, le RSB est connu dans chaque bande. Le RSB dans chaque bande sera modifié suivant sa valeur et son emplacement. L'objectif de ces modifications est de soustraire le plus de bruit dans les bandes où nous trouvons un RSB faible. Pour chaque trame m , nous vérifions le niveau du RSB dans les treize dernières bandes ($5 \leq i < N_b$). Ensuite, si nous trouvons que nous avons six bandes ou plus, dans les treize dernières bandes, ayant un RSB inférieur à 4 dB, c'est-à-dire un niveau élevé du bruit dans cette trame, nous procédons à d'autres modifications.

Celles-ci consiste à mettre à 2 dB toutes les bandes qui ont un RSB inférieur ou égal à 4 dB. Toutes les bandes qui ont un RSB compris entre 4 et 6 dB sont mis à 4 dB et les bandes entre 6 et 8 à 6 dB. Le but de ces modifications est de soustraire le plus de bruit dans cette trame, car elle contient une quantité élevée de bruit.

Cependant, si le « flag » déclenché par le DAV est mis à 1 signifiant la présence d'une trame de bruit, alors nous mettons toutes les bandes à 1 dB et ainsi la quantité de bruit soustraite serait constante sur tout le spectre. Ceci a pour but de décaler le spectre du bruit vers le bas de 13 dB. Le choix de ce nombre est flexible et ce, selon l'application. Dans le cas de la parole, nous n'enlevons pas tout le bruit pour ne pas créer un silence total. Car, si nous enlevons plus que 13 dB, le signal deviendrait synthétique et désagréable à l'écoute (période silencieuse).

La figure 3.10 décrit les spectres d'amplitude de deux signaux de parole dans la situation où le « flag » est à 1 (trames de bruit):

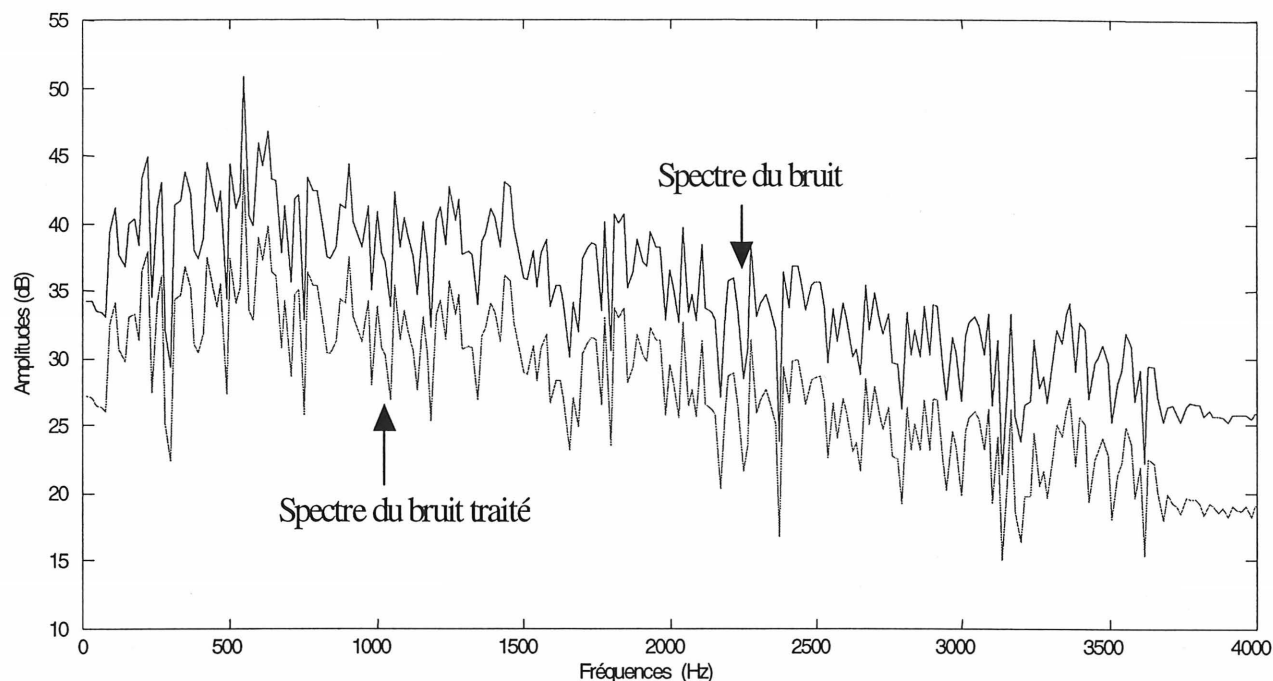


Figure 3.10 Deux spectres de bruit avant et après traitement.

Finalement, quelque soit les conditions, toutes les valeurs inférieures à 2 dB sont mises à 1dB.

3.7 Estimation du Bruit

Pour estimer le bruit de l'environnement du signal de parole, nous initialisons le bruit pour les quatre premières trames avec l'énergie du canal, c'est-à-dire :

$$E_b(m, i) = E_c(m, i) \quad 1 \leq m \leq 4 \quad \text{et} \quad 0 \leq i < N_b \quad 3-10$$

Ensuite, à chaque fois que le « flag » de DAV passe à 1, nous faisons la mise à jour du bruit dans chaque canal, ce qui peut s'exprimer comme suit :

$$E_b(m+1, i) = \max\{E_{\min}, \alpha_b E_b(m, i) + (1 - \alpha_b) E_c(m, i)\} \quad 0 \leq i < N_b \quad (3-11)$$

avec $E_{\min} = 0.1$ et $\alpha_b = 0.9$.

où E_{\min} est l'énergie minimale permise pour chaque canal et α_b est le facteur de lissage du bruit par bande.

Donc, pour éviter tout changement rapide de l'énergie estimée du bruit, nous utilisons un facteur de lissage de 0.1. C'est-à-dire que nous prenons 0.1 de la trame actuelle et 0.9 de celle du passé. Ainsi, le bruit est lissé sur 10 trames du passé. Ce calcul d'énergie est entièrement tiré de l'article [11] de Ramabadran et al. Aucune raison de le changer ne semble se présenter.

3.8 Calcul du gain γ_g par bande

Toutes les étapes précédentes sont réalisées pour que nous puissions calculer le meilleur gain pour chaque bande critique. Le calcul de ce gain se fait en trois étapes. La première étape consiste à calculer le gain en général tel que donnée dans l'équation suivante :

$$\gamma_g = \max \left\{ \gamma_{\min}, \min \left\{ 0, -10 * \log_{10} \left(\frac{\sum_{i=5}^{N_b-5} E_b(i)}{N_b - 10} \right) \right\} \right\} \quad (3-12)$$

avec $\gamma_{\min} = -13$ dB et $-13 \text{ dB} \leq \gamma_g \leq 0$ dB.

L'équation 3-12 se base sur l'énergie du bruit. Cette étape permet de calculer la quantité de bruit à soustraire du RSB. La limitation du γ_g à -13 dB est un choix que nous avons pris après plusieurs itérations.

La deuxième étape consiste à calculer le gain du canal en dB, γ_{dB} ; donné par :

$$\gamma_{dB}[i] = RSB[i] + \gamma_g \quad (3-13)$$

Par la suite, nous soustrairons la valeur calculée de γ_g du RSB. Ceci nous permet de calculer le gain pour chaque bande critique. Donc, pour chacune des 18 bandes critiques, nous avons un gain différent en fonction de la valeur du RSB.

La troisième étape consiste à convertir le γ_{dB} en gain linéaire, γ_c ; comme suit :

$$\gamma_c[i] = \min \left\{ 1, 10^{\frac{\gamma_{dB}(i)}{20}} \right\} \quad 0 \leq i < N_b \quad (3-14)$$

où $0 \leq \gamma_c \leq 1$. Car si $\gamma_{dB} > 0$, γ_c reste toujours à 1 ($10^x > 0$ si $x > 0$).

3.9 Filtrage dans le domaine fréquentiel

Tous les calculs précédents avaient pour objectif de trouver le meilleur gain qui multiplie le spectre du signal bruité. Après avoir calculer le meilleur gain par canal, γ_c , nous l'avons multiplié par le spectre d'amplitude du signal bruité, $G(k)$.

$$H(k) = \begin{cases} \gamma_c[i]G(k) & 1 \leq i < N_b - 1 \text{ et } f_l[i] \leq k < f_n[i] \\ G(k) & i = 0 \text{ et } i = N_b - 1, f_l[i] \leq k < f_n[i] \text{ et } k = L/2 \end{cases} \quad (3-15)$$

où $f_n[0]$ et $f_n[N_b-1]$ sont respectivement le nombre de raies dans la première et la dernière bande. $L/2$ correspond à la dernière fréquence dans le spectre d'amplitude, c'est-à-dire que les basses fréquences et les hautes fréquences du spectre d'amplitude ne sont pas modifiées, car elles vont être filtrées par le canal de transmission 300-3400 Hz.

3.10 IDFT, fenêtrage, chevauchement et addition

En faisant la IDFT, nous ramenons le signal du domaine fréquentiel à celui du domaine du temps. La présence d'une racine carrée à la fenêtre d'analyse, exige de fenêtrer le signal à la synthèse pour multiplier ce signal une deuxième fois par une racine carrée. Alors, la fenêtre de reconstruction est aussi une fenêtre sinusoïdale (racine carrée de la fenêtre de Hanning). Pour que l'addition et la multiplication de ces deux fenêtres donne 1, nous devons multiplier le signal par cette fenêtre et l'additionner sur un chevauchement de 50%, 128 échantillons.

3.10.1 Transformée de Fourier Discrète Inverse (IDFT)

Avant de calculer la DFT inverse, nous calculons les parties réelles et imaginaires de la transformée définies par :

$$R(k) = \sum_{k=0}^{L/2} H[k] \cos(\varphi[k]) \quad (3-16)$$

$$I(k) = \sum_{k=0}^{L/2} H[k] \sin(\varphi[k]) \quad (3-17)$$

La fonction IDFT sort un signal réel.

La définition d'une DFT inverse de L points complexes est donnée par:

$$x(n) = \frac{1}{\sqrt{L}} \sum_{k=0}^{L-1} X(k) e^{(j \frac{2\pi nk}{L})} \quad (3-18)$$

$$x(n) = \frac{1}{\sqrt{L}} \sum_{k=0}^{L-1} (R(k) + jI(k)) e^{j\frac{2\pi nk}{L}} \quad (3-19)$$

où $n = 0, 1, 2, \dots, L-1$

Donc la partie réelle $r(n)$ est égale à :

$$r(n) = \frac{1}{\sqrt{L}} \sum_{k=0}^{L-1} R(k) \cos\left(\frac{2n\pi k}{L}\right) - I(k) \sin\left(\frac{2n\pi k}{L}\right) \quad (3-20)$$

et la partie imaginaire $i(n)$:

$$i(n) = \frac{1}{\sqrt{L}} \sum_{k=0}^{L-1} I(k) \cos\left(\frac{2n\pi k}{L}\right) + R(k) \sin\left(\frac{2n\pi k}{L}\right) \quad (3-21)$$

Pour un signal $x(n)$ réel, nous avons les propriétés suivantes pour $R(k)$ et $I(k)$:

$$R(k) = R(L-k) \quad R(k) \text{ est la partie réelle de } X(k)$$

$$I(k) = -I(L-k) \quad I(k) \text{ est la partie imaginaire de } X(k)$$

$$\text{où } k = 0, 1, 2, \dots, L/2$$

Étant donné ces symétries, la partie imaginaire $i(n) = 0$ et la partie réelle devient:

$$r(n) = \frac{2}{\sqrt{L}} \sum_{k=1}^{L/2-1} R(k) \cos\left(\frac{2n\pi k}{L}\right) - \frac{2}{\sqrt{L}} \sum_{k=1}^{L/2-1} I(k) \sin\left(\frac{2n\pi k}{L}\right) + \frac{1}{\sqrt{L}} (R(0) + (-1)^n R(L/2)) \quad (3-22)$$

où $n = 0, 1, 2, \dots, L-1$.

Donc, la sortie de la fonction IDFT est $r(n)$ qui est le signal de parole dans le domaine du temps.

3.10.2 Fenêtre sinusoïdale de reconstruction

Le calcul final du signal dans le domaine du temps se fait comme suivant :

$$y(i) = w(i+128) * r_passé(i) + w(i) * r(i) \quad 0 \leq i < 128 \quad (3-23)$$

$w(i)$ est la fenêtre sinusoïdale de reconstruction et $r(i)$ est le signal dans le domaine du temps.

En plus, $r_passé$ présente le signal $r(i)$ de la trame $m-1$:

$$r_passé(i) = r(i+128) \quad 0 \leq i < 128 \quad (3-24)$$

Rendu à cette étape, nous terminons l'analyse totale du schéma-bloc de la figure 3.1. Les résultats de cet algorithme seront présentés au chapitre 4.

Chapitre 4

Résultats

4.1 Introduction

La troisième étape de notre travail consiste à présenter l'analyse globale des résultats de tests obtenus dépendamment des attentes fixées au départ et de la revue bibliographique. Un test d'écoute a été réalisé sur six fichiers de parole (3 signaux pour les voix d'hommes et 3 signaux pour les voix de femmes) sur lesquels nous avons ajouté six genres de bruit différents, à deux niveaux (10 dB et 5 dB). Avec ce test, nous avons comparé nos résultats à ceux du réducteur de bruit réalisé par Ramabadran et al, [11].

Puisque nous avons six genres de bruit à deux niveaux, alors nous avons fait un test d'écoute sur douze fichiers de parole.

4.2 Méthode du test

Les bruits qui ont été ajoutés aux signaux de parole propre sont les suivants : le bruit blanc stationnaire, le vent qui siffle, le vent accompagné d'un bruit de véhicules variable, le bruit de bureau avec du bruit de bafouille, le bruit de la rue comprenant celui de véhicules (bruit variable) et un genre de bruit très variable qui englobe du bruit de bafouille, voitures et vent.

La méthode du test est d'écouter le signal de synthèse de l'algorithme proposé (A), celui proposé par Ramabadran et al (B) et de voter par préférence : A, B ou égale.

4.3 Résultats de tests d'écoute

Dans les tableaux 4-1 à 4-3, nous présentons les pourcentages de votes qui ont été réalisés auprès de 13 auditeurs :

Le tableau 4-1 regroupe les résultats de test pour les signaux de parole avec une voix d'homme à deux niveaux de bruit.

Tableau 4-1 Pourcentage de votes pour des signaux de parole avec une voix d'homme à deux niveaux de bruit. Le nombre d'auditeurs est de treize.

Genre du bruit	Voix d'homme, 10 dB			Voix d'homme, 5 dB		
	<i>L'algorithme proposé</i>	<i>Égalités</i>	<i>Algorithme de Ramabadran et al</i>	<i>L'algorithme proposé</i>	<i>Égalités</i>	<i>Algorithme de Ramabadran et al</i>
<i>Bruit blanc stationnaire</i>	77%	8%	15%	31%	31%	39%
<i>Vent avec bruit de véhicules</i>	39%	31%	31%	15%	46%	39%
<i>Vent qui siffle</i>	39%	23%	39%	23%	46%	31%

Nous remarquons que les signaux de parole bruités à 10 dB sont mieux traités par l'algorithme proposé comparativement à celui de Ramabadran et al. Par contre, l'algorithme de Ramabadran et al traite mieux les signaux avec un niveau de bruit de 5 dB. À notre avis, le problème se confine autour du DAV [18] qui, dans certaines conditions, ne se déclenche pas adéquatement. Donc, des améliorations éventuelles devront se faire au niveau du DAV.

Le tableau 4-2 présente des signaux de parole avec une voix de femme à deux niveaux de bruit.

Tableau 4-2 Pourcentage de votes pour des signaux de parole avec une voix de femme à deux niveaux de bruits. Le nombre des auditeurs est de treize.

Genre du bruit	Voix de femme, 10 dB			Voix de femme, 5 dB		
	<i>L'algorithme proposé</i>	<i>Égalités</i>	<i>Algorithme de Ramabadran et al</i>	<i>L'algorithme proposé</i>	<i>Égalités</i>	<i>Algorithme de Ramabadran et al</i>
<i>Bruit blanc avec bruit de bafouille</i>	85%	8%	8%	54%	31%	15%
<i>Bruit de bafouille avec bruit du vent et de voitures</i>	39%	39%	23%	15%	8%	77%
<i>Bruit de la rue et des voitures</i>	85%	0%	15%	69%	15%	15%

Nous remarquons que l'algorithme proposé réagit mieux avec les voix de femmes. Même à 5 dB, ce tableau présente de très bons résultats, sauf à 5 dB où le bruit de bafouille, du vent et de voitures se trouvent en même temps ceci peut être dû au DAV. Si nous faisons la comparaison dans le cas de 5 dB entre les voix d'hommes et celles de femmes, nous remarquons que l'algorithme proposé réagit mieux dans la deuxième cas.

Une moyenne arithmétique englobant l'ensemble des tests, pour les deux niveaux de bruit, est présentée dans le tableau 4.3. En somme, les résultats sont satisfaisants tout en gardant à l'esprit que particulièrement dans le cas d'un RSB plus faible, des améliorations notables peuvent être réalisées :

- 1- Changement au niveau du DAV pour qu'il soit optimum dans tous les cas de bruits.
- 2- Améliorations de gain par bande.
- 3- Trouver une autre fenêtre que celle proposée, ceci est pour améliorer le chevauchement.

Tableau 4-3 Moyenne générale des pourcentages de votes pour les signaux avec les voix d'hommes et de femmes à deux niveaux de bruit.

	10 dB			5 dB		
	<i>L'algorithme proposé</i>	<i>Égalités</i>	<i>Algorithme de Ramabadran et al</i>	<i>L'algorithme proposé</i>	<i>Égalités</i>	<i>Algorithme de Ramabadran et al</i>
Moyenne générale	60%	18%	22%	35%	29%	36%

Les résultats sont également présentés sous forme d'histogrammes (figures 4.1 et 4.2), illustrant respectivement les compilations des moyennes obtenues pour les signaux dont le RSB est de 10 dB et 5 dB.

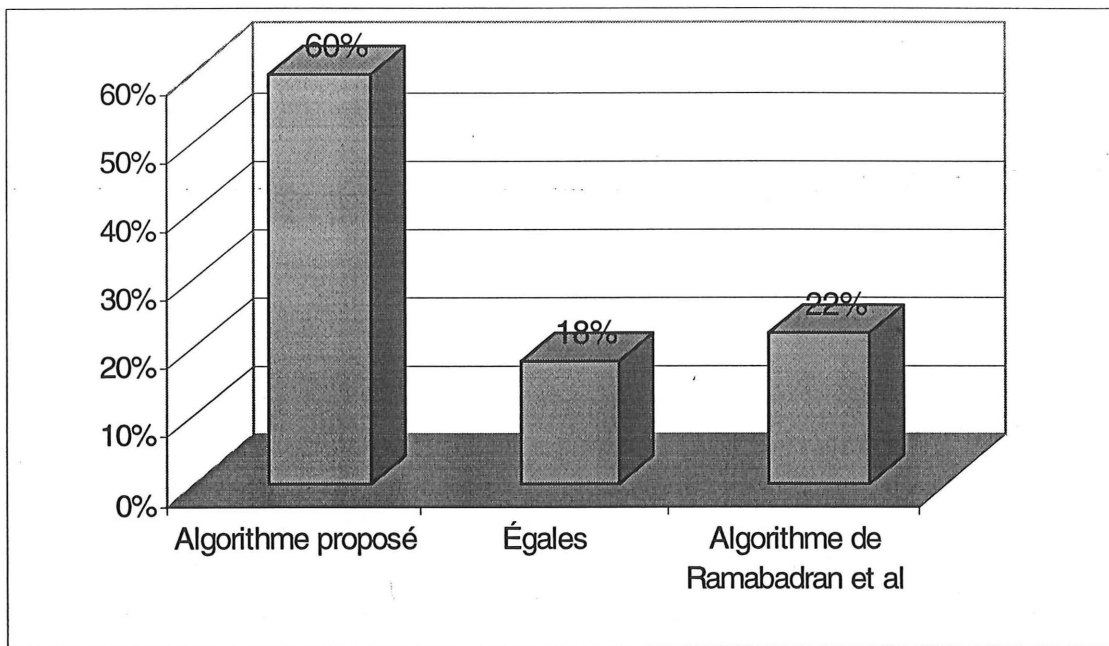


Figure 4.1 Préférence subjective des réducteurs de bruit pour un RSB de 10 dB.

À 10 dB, l'algorithme proposé recueille 3 fois plus de votes que celui de Ramabadran et al, ce qui constitue un écart très significatif. Pour un RSB de 5 dB, les résultats sont similaires et sont attribuables essentiellement à la complexité du travail effectué à ce niveau et le problème engendré par le déclenchement erroné du DAV.

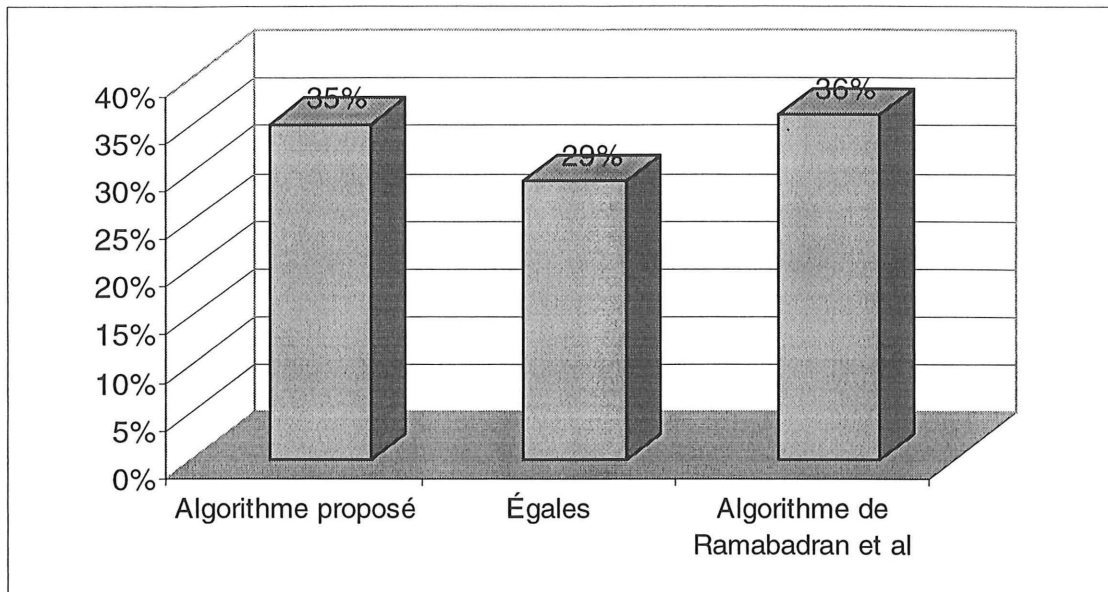


Figure 4.2 Préférence subjective des réducteurs de bruit pour un RSB de 5 dB.

À la figure 4.3, nous présentons un signal de parole bruité avec un niveau de bruit de 10 dB (signal d'analyse) et le signal débruité avec l'algorithme proposé (signal de synthèse).

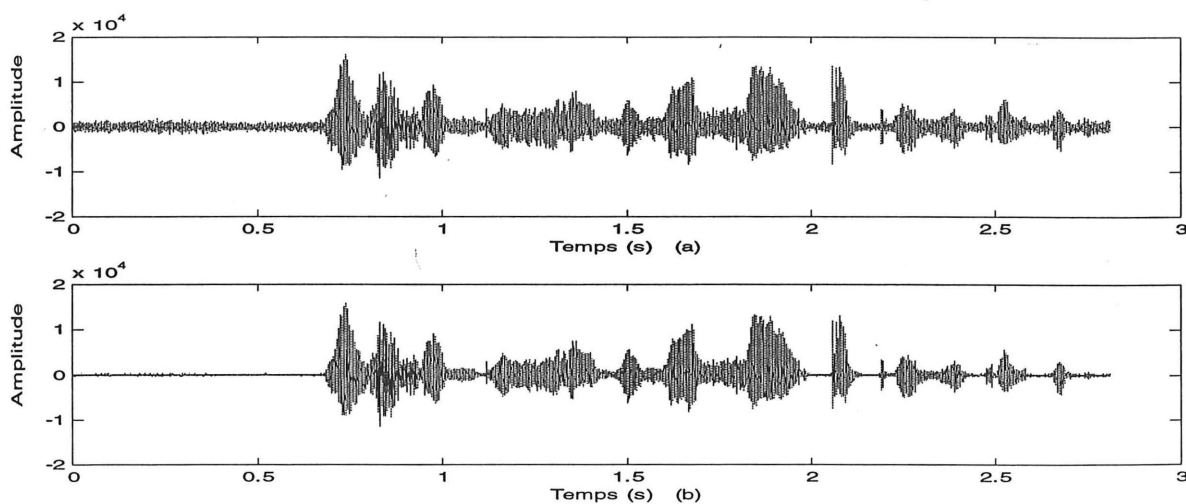


Figure 4.3 Réduction de bruit. (a) Signal bruité avec RSB de 10 dB. (b) Signal traité.

La mise en relief de l'effet de la réduction du bruit sur le spectre d'amplitude est illustré à la figure 4.4 pour une portion voisée du signal sur lequel une DFT a été effectuée.

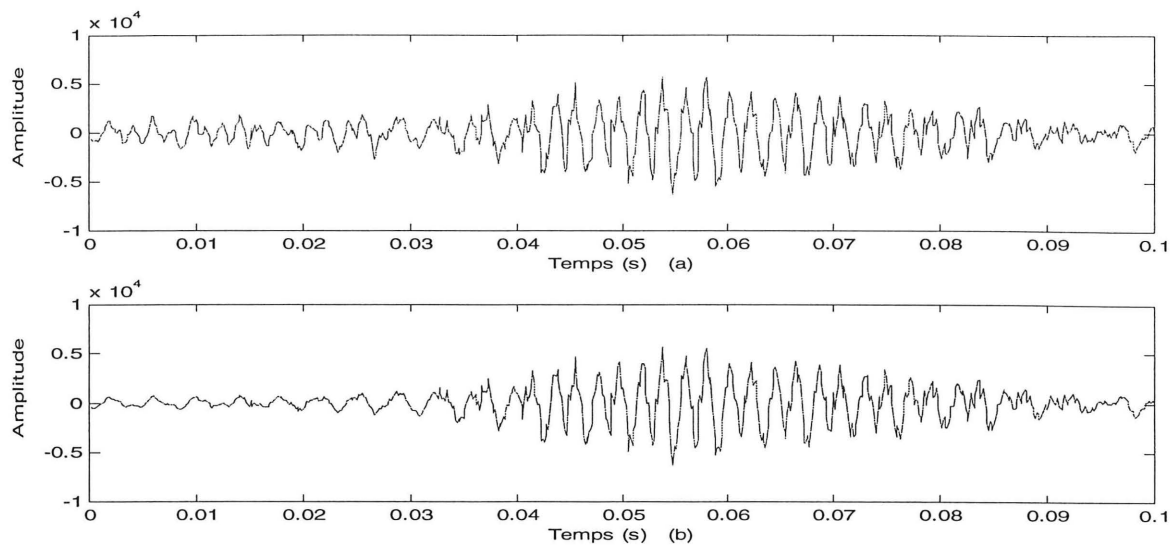


Figure 4.4 Parties voisées des signaux de parole de la figure 4.3.

À partir de ces signaux, nous avons effectué une DFT de 1024 points (figure 4.5).

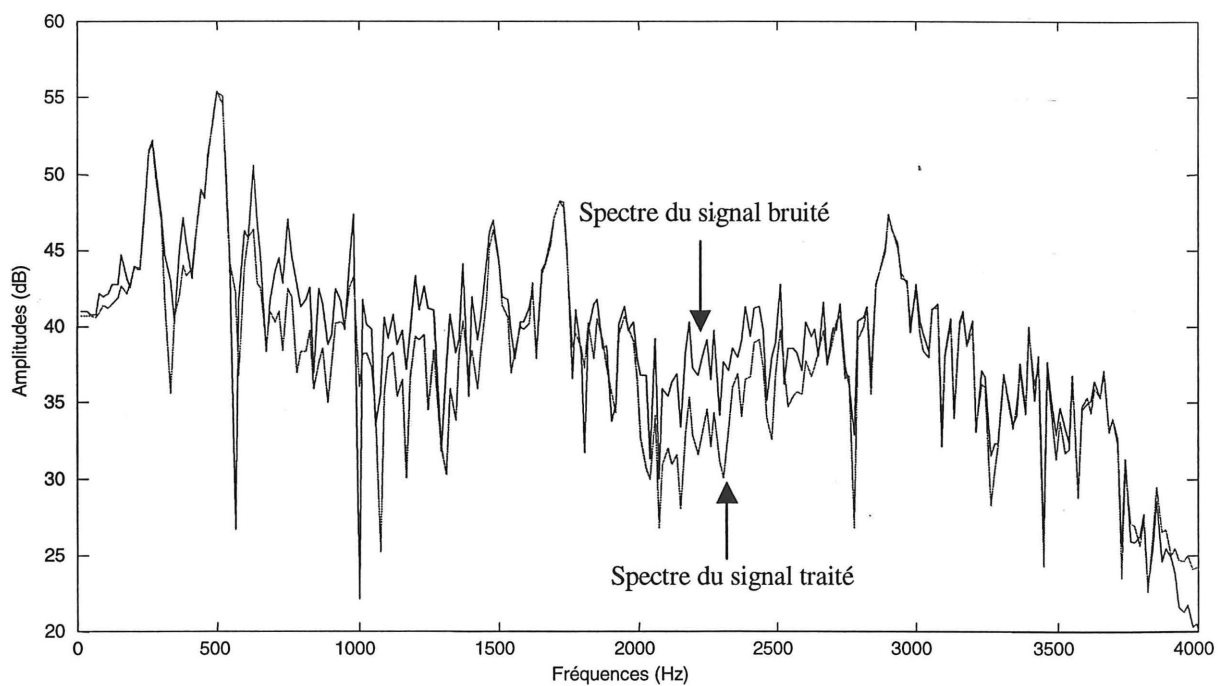


Figure 4.5 Spectres d'amplitude des deux signaux de la figure 4.4.

Le spectre du signal débruité ne change pas dans les endroits où l'énergie est élevée. Par contre, dans les endroits où l'énergie est faible, le spectre du signal bruité est tiré en bas de quelques dB,

Réduction de bruit
dans un signal vocal pour
les bandes étroite et élargie

tout dépendant de niveau du RSB. Donc, nous creusons le spectre du signal bruité dans les endroits où l'énergie est faible.

Par la suite, dans le but de montrer la différence dans les spectres d'amplitude, nous faisons ressortir une partie des signaux de la figure 4.3 qui ne contient que du bruit.

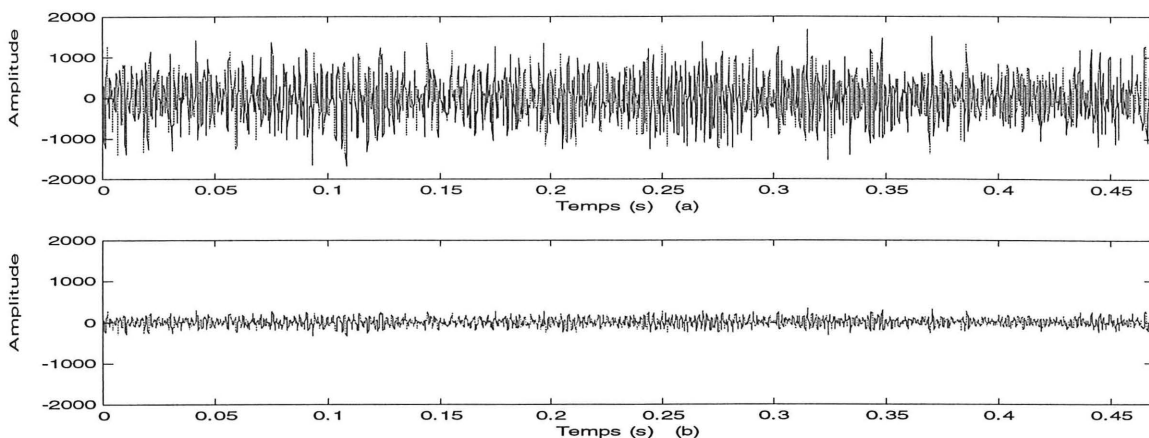


Figure 4.6 Réduction de bruit. (a) Trames du bruit seulement avec un RSB de 5 dB. (b) Signal traité.

Les spectres de ces deux signaux sont présentés à la figure 4.7 :

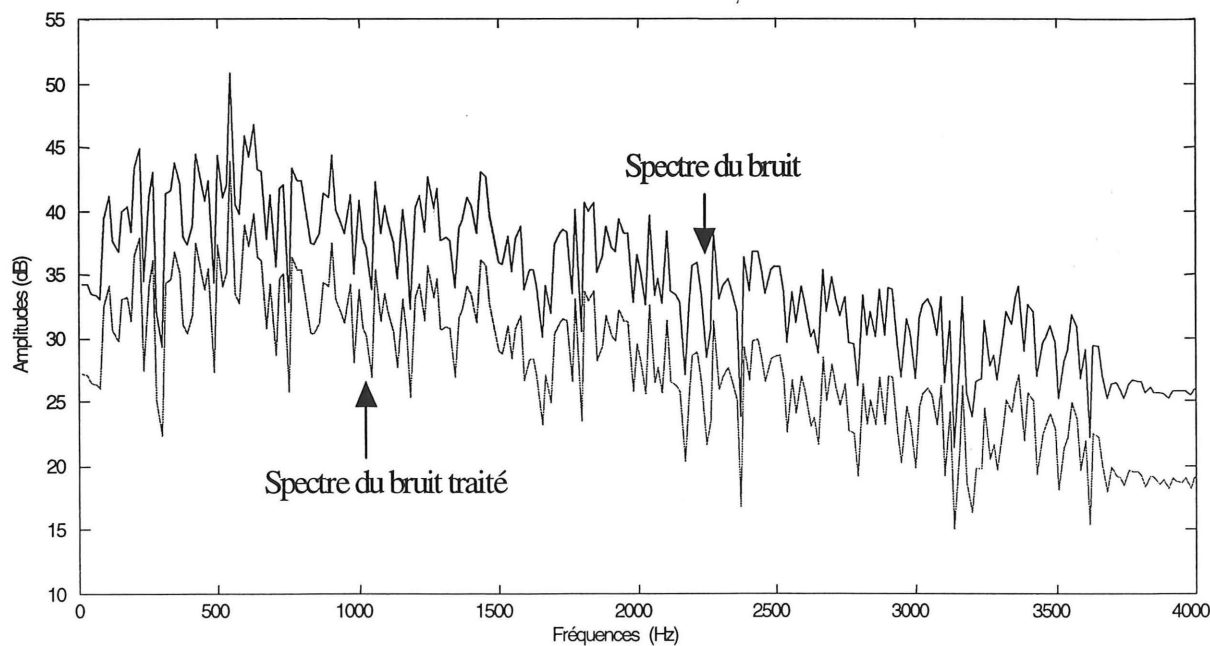


Figure 4.7 Spectres d'amplitude des deux signaux présentés à la figure 4.6.

Ces spectres sont calculés sur 512 points. Le choix du 13 dB représente le meilleur compromis afin de garder l'allure du spectre du bruit et de ne pas enlever tout le bruit afin de garder une allure naturelle à la communication. Ceci est appelé : bruit de confort.

Il faut sans doute noter que le réducteur du bruit ne doit pas modifier un signal de parole propre. C'est-à-dire que si le signal de parole n'est pas bruité, le réducteur du bruit ne doit pas déformer ou changer le spectre du signal original. La figure 4.8 (a) illustre le signal de parole propre et la figure 4.8 (b), le signal traité par le réducteur du bruit.

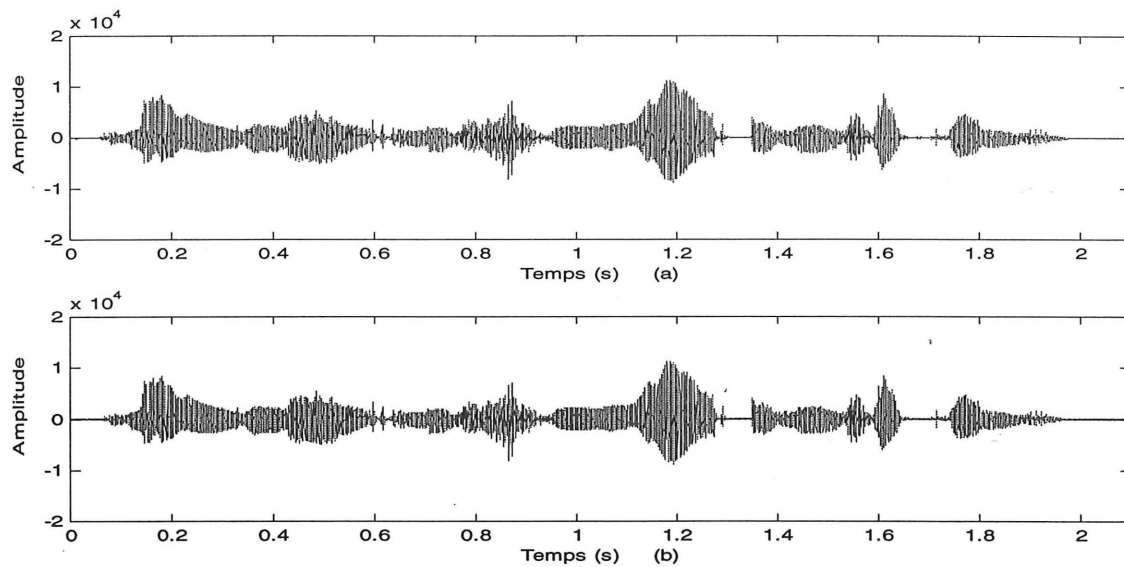


Figure 4.8 (a) Signal de parole propre original. (b) Signal traité avec l'algorithme proposé.

Pour mieux voir les spectres de ces signaux, nous présenterons à la figure 4.9 une partie des signaux de la figure 4.8.

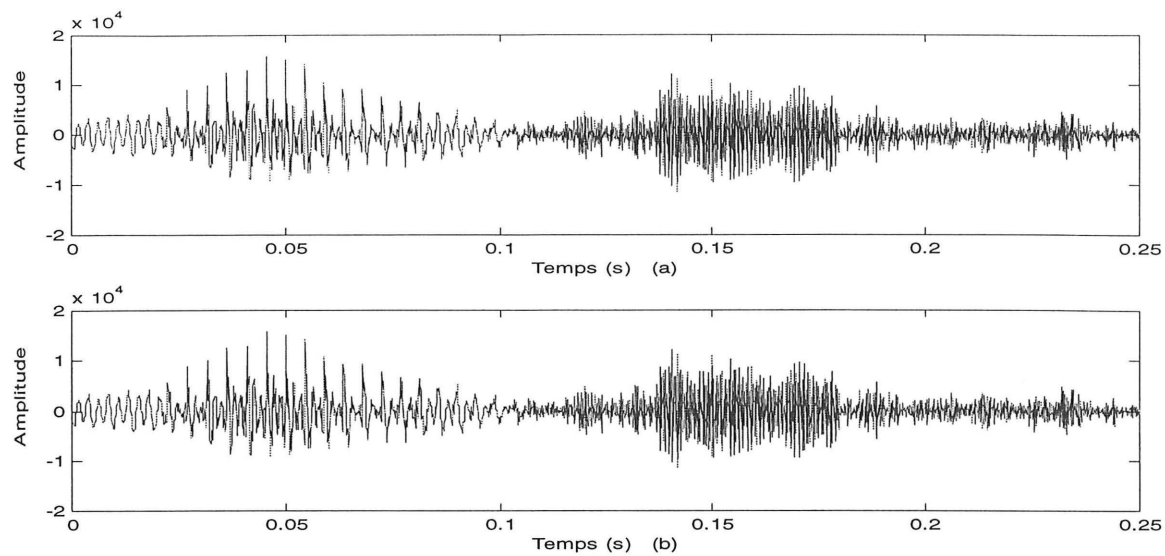


Figure 4.9 (a) Partie de signal de la figure 4.8 (a). (b) Partie de signal de la figure 4.8 (b).

Nous présentons à la figure 4.10 les spectres de ces deux signaux qui ont été faits par une DFT de 256 points.

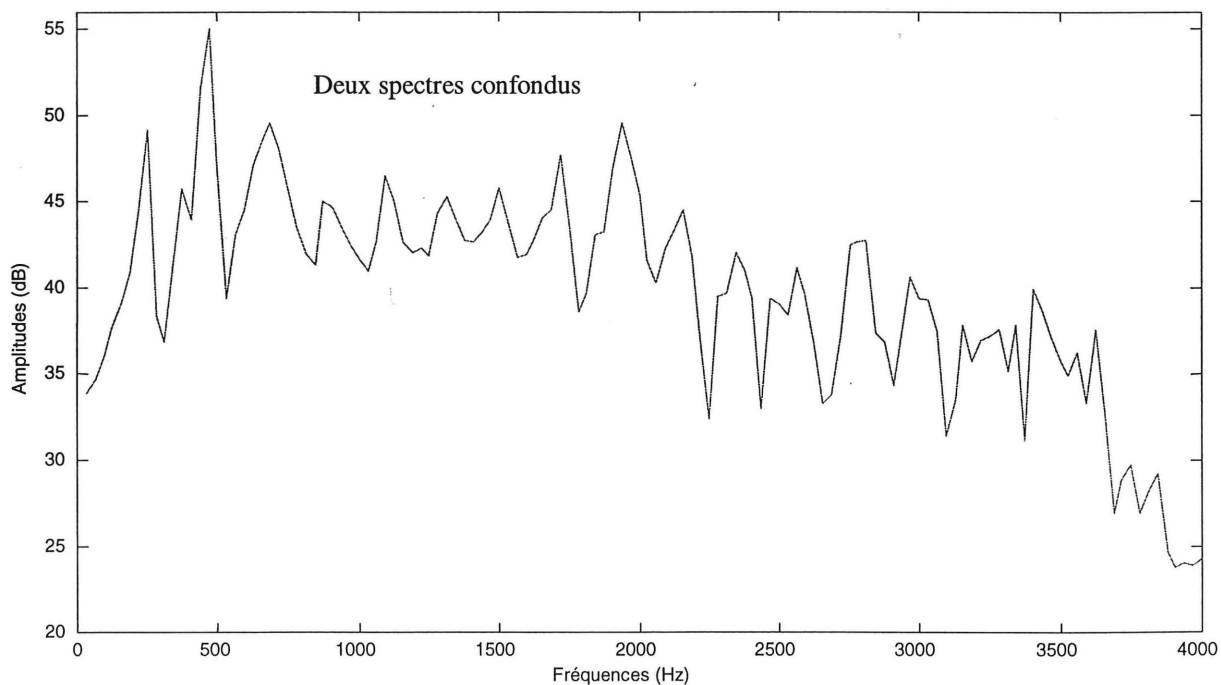


Figure 4.10 Deux spectres confondus des deux signaux de la figure 4.9.

La figure 4.10 illustre les deux spectres présentées aux figures 4.9 (a) et 4.9 (b). Les deux spectres sont confondus. Ceux-ci montrent bien que l'algorithme proposé ne modifie pas le spectre d'un signal propre. Ce point présente une amélioration remarquable par rapport au réducteur de bruit proposé par Ramabadran et al, car dans certains cas où le signal à traiter est propre, l'algorithme de Ramabadran et al atténue les hautes fréquences; ce qui crée une différence entre le signal original propre et le signal débruité. Par contre, l'algorithme proposé évite ce genre de problèmes.

Pour répondre à la forte demande du marché actuel, le chapitre suivant fait l'objet de travaux qui ont permis d'étendre le champ de travail en présentant un algorithme pour une bande élargie de 8 kHz.

Chapitre 5

Réducteur de bruit pour la large bande

5.1 Introduction

Plusieurs systèmes de communication prennent un tournant vers l'utilisation d'une bande élargie pour répondre aux exigences de la qualité dans la transmission vocale. Avec l'avènement de la téléconférence, de la transmission de voix par internet et d'autres moyens de télécommunication, la qualité de la voix est devenue plus importante que jamais. Dans les applications de téléphonies actuels, le signal d'entrée est filtré entre 200 et 3400 Hz. Ceci affecte son audition naturel (manque de basses fréquences) et son intelligibilité (manque de hautes fréquences). Une bande élargie de 50-7400 Hz est essentielle pour garantir une qualité de communication élevée.

Ce besoin tangible nous a poussé à réaliser un algorithme qui réduit le bruit pour des signaux de parole échantillonnés à 16 kHz. Donc, nous sommes en présence d'une bande qui représente le double de la bande téléphonique de 4 kHz.

5.2 Détails de l'algorithme

Le principe de fonctionnement général de l'algorithme à bande élargie est le même qu'au chapitre 3. Le schéma-bloc du réducteur de bruit pour la bande élargie est le même que celui pour la bande étroite (figure 3.1). Cependant, certaines modifications ont été faites à l'intérieur

de chaque bloc. Dans ce chapitre nous ne rentrerons pas en détail dans chaque bloc, mais nous présenterons seulement les modifications apportées au schéma-bloc du chapitre 3 pour le signal vocal à large bande.

5.3 Fenêtrage et DFT

Il est à noter que la longueur de la trame reste de 16 ms. Puisque le signal d'analyse est échantillonné à 16 kHz, alors dans chaque trame de 16 ms, nous trouvons 256 échantillons. Nous avons utilisé la même fenêtre sinusoïdale décrite au chapitre 3. Mais, la longueur de cette fenêtre est de 512. Le chevauchement reste à 50% et la DFT se calcule sur 512 points.

5.4 Estimation de l'énergie du canal

Avant d'estimer l'énergie du canal pour chaque trame m , nous décomposons le spectre d'amplitude en 22 bandes critiques $B(i)$ avec $0 \leq i < 22$. Si la fréquence d'échantillonnage double, le nombre de bandes critiques ne double pas. C'est-à-dire que les bandes critiques de l'algorithme à large bande n'ont pas passé de 18, bandes calculées dans le cas de la bande étroite, à 36. Mais, seulement à 22 bandes, car le nombre des raies a augmenté dans les bandes qui contiennent les hautes fréquences. Le tableau 5-1 montre le nombre de raies qui existent dans chaque bande critique $B(i)$:

Tableau 5-1 Bandes critiques pour la bande élargie.

Bande critique $i^{ème}$ nombre,	Les raies du DFT		Fréquences réelles [Hz]
	Intervalles	Nombre des raies	
1	1-3	3	0-94
2	4-6	3	94-187
3	7-10	4	187-312
4	11-13	3	312-406
5	14-16	3	406-500
6	17-20	4	500-625
7	21-25	5	625-781
8	26-29	4	781-906
9	30-35	6	906-1094
10	36-41	6	1094-1281
11	42-47	6	1281-1469
12	48-55	8	1469-1719
13	56-64	9	1719-2000
14	65-74	10	2000-2312
15	75-86	12	2312-2687
16	87-100	14	2687-3125
17	101-118	18	3125-3687
18	119-140	22	3687-4375
19	141-169	29	4375-5281
20	170-204	35	5281-6375
21	205-246	42	6375-7687
22	247-256	10	7687-8000

Le calcul de l'énergie dans chaque bande critique $B(i)$ dans la première trame se fait comme suit :

$$E_c(m, i) = \sum_{i=0}^{N_b-1} \sum_{k=0}^{f_n[i]} \frac{(G[f_L[i] + k])^2}{f_n[i]} \quad \text{avec } m=1 \text{ et } 0 \leq N < N_b \quad (5-1)$$

où N_b est le nombre de bandes $B(i)$, $N_b = 22$.

$G(k)$ est le spectre d'amplitude.

$f_L[i]$ est la partie inférieure de l'intervalle :

$f_L[i] = \{0, 3, 6, 9, 12, 16, 20, 24, 29, 34, 40, 47, 55, 64, 74, 86, 100, 118, 140, 169, 204, 246\}$;

$f_n[i]$ est le nombre d'échantillons dans chaque bande :

$f_n[i] = \{3, 3, 3, 3, 4, 4, 4, 5, 5, 6, 7, 8, 9, 10, 12, 14, 18, 22, 29, 35, 42, 10\}$.

L'énergie est lissée par bande par un facteur de 0.45, telle que présenté à la section 3.3.

5.5 Détecteur Parole / Bruit, (DAV)

Les mêmes principes que ceux du DAV pour la bande étroite présentée au chapitre 3 (section 3.4) ont été utilisés pour réaliser le DAV pour la large bande. Toutefois, l'algorithme du DAV pour la bande étroite a été étendu pour la large bande.

Ce travail a été réalisé par un chercheur de notre groupe de recherche [19], mais nous l'avons intégré dans notre code sans le modifier. Donc, c'est un long travail de recherche qui ne rentre pas dans la présente étude.

5.6 Estimation du RSB par canal

Cette étape de calcul est identique à la section 3.5. Sauf que, $N_b = 22$ au lieu de 18.

5.7 Modification du RSB par canal

À partir des sections 3.5 et 5.6 nous constatons que le RSB est connu dans chaque bande. La valeur de chaque RSB est modifiée suivant son niveau et son emplacement dans les bandes critiques.

Pour chaque trame m , nous vérifions le niveau du RSB dans les 18 dernières bandes ($3 \leq i < N_b - 1$). Ensuite, si nous trouvons que nous avons 8 bandes ou plus dans ces 18 bandes qui ont un RSB inférieur à 5 dB, nous procédons aux modifications suivantes :

- 1- Toutes les bandes qui ont un RSB inférieur à 5 dB sont ramenées à 2 dB.
- 2- Si le RSB dans chaque bande est compris entre 5 et 8 dB, nous le ramenons à 4 dB.

Si nous trouvons, dans une trame donnée, plus de 8 bandes ayant un RSB inférieur à 5 dB, alors nous considérons que c'est une trame avec un niveau de bruit élevé. Donc, nous devons diminuer la valeur de RSB pour que la soustraction de bruit soit plus élevée.

Ensuite, si le « flag » du DAV passe de 0 à 1 (présence d'une trame de bruit), toutes les bandes seront mis à 1 dB. Enfin, indépendamment des conditions, toutes les valeurs inférieures à 2 dB seront mises à 1 dB.

5.8 Calcul du gain γ_c par bande

Le calcul du gain par bande s'opère identiquement à celui de la section 3.8.

5.9 Filtrage dans le domaine fréquentiel

D'une façon similaire à la méthode décrite dans le chapitre 3, nous appliquons le gain par bande γ_c sur le spectre d'amplitude du signal bruité, $G(k)$.

$$H(k) = \begin{cases} \gamma_c[i]G(k) & 1 \leq i < N_b - 1 \text{ et } f_l[i] \leq k < f_n[i] \\ G(k) & i = 0 \text{ et } i = N_b - 1, f_l[i] \leq k < f_n[i] \text{ et } k = L/2 \end{cases} \quad (5-2)$$

La deuxième partie de l'équation 5-2 montre que la première et la dernière bande ne seront pas modifiées puisqu'elles vont être filtrées par le canal de transmission. $L/2$ présente la dernière fréquence dans le spectre d'amplitude.

5.10 Estimation du Bruit

À chaque fois que le « flag » passe à 1, nous faisons la mise à jour du bruit comme mentionné à la section 3.7.

5.11 IDFT, fenêtrage , chevauchement et addition (overlap Add)

Cette méthode ne présente aucun changement par rapport à celle du chapitre 3 à part le fait que la longueur de la fenêtre sinusoïdale de reconstruction devient de 512 au lieu de 256.

5.12 Résultats

Dans un premier temps, nous montrons des signaux de parole échantillonnés à 16 khz traités par cet algorithme et leurs spectres d'amplitude. Ceci nous permet de les analyser et de voir l'effet de la soustraction spectrale sur ces spectres d'amplitudes.

La première figure dans ce chapitre décrit deux signaux de parole dont la figure 5.1 (a) illustre le signal de parole bruité avec un RSB de 5 dB et celle 5.1 (b) correspond au signal traité par l'algorithme proposé.

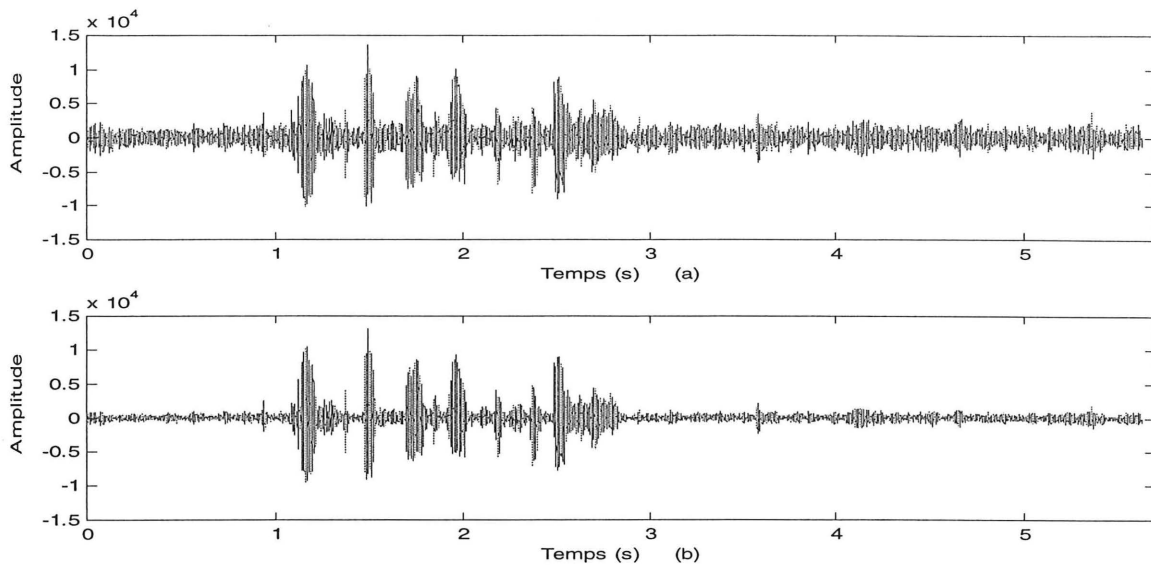


Figure 5.1 Réduction de bruit. (a) Signal de parole bruité avec un RSB de 5 dB. (b) Signal débruité par l'algorithme proposé.

Les trames qui se trouvent après la troisième seconde sont des trames de bruit. Mais, ce bruit est un bruit de bafouille (babble noise). Ce genre de bruit est très difficile à traiter, car ces trames ressemblent à la voix. Donc, le DAV ne déclenche pas. La figure 5.2 illustre deux petites parties voisées des signaux à la figure 5.1.

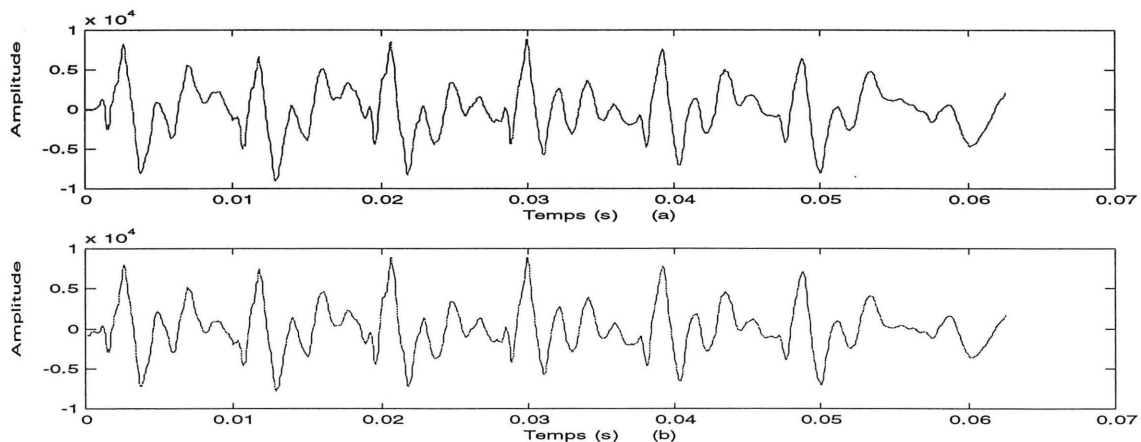


Figure 5.2 (a) Signal de parole voisé bruité avec un RSB de 5 dB. (b) Signal traité.

Pour voir l'effet du réducteur de bruit sur le spectre d'amplitude du signal bruité nous présentons les spectres d'amplitudes de ces deux signaux voisés.

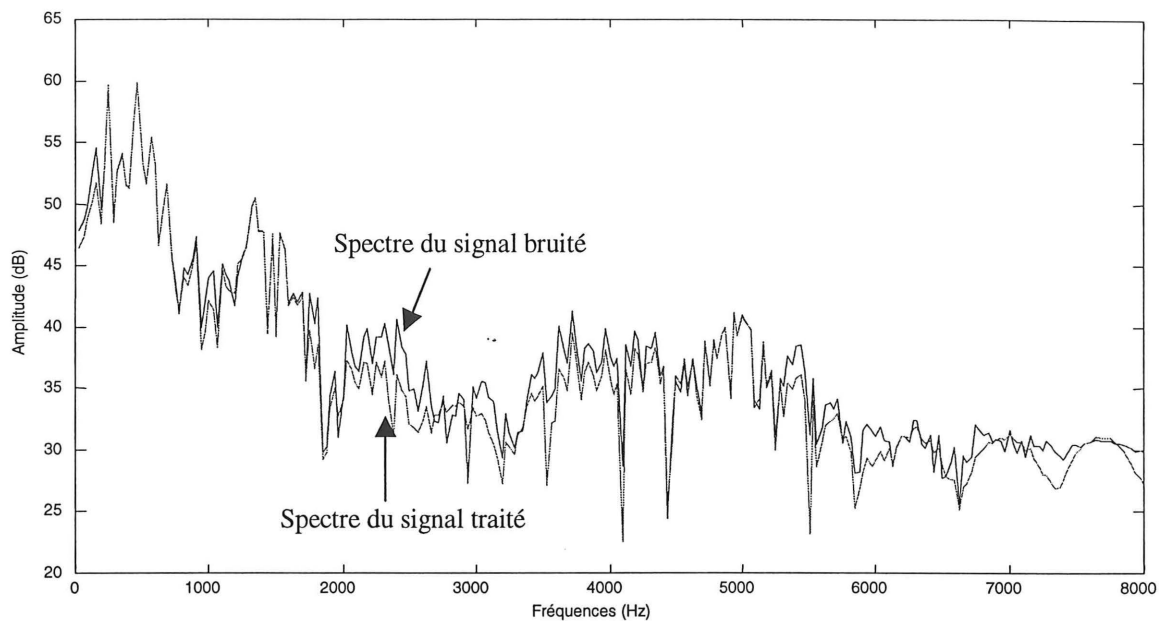


Figure 5.3 Spectres d'amplitude des deux signaux voisés de la figure 5.2.

Puisque nous traitons un signal de parole voisé, nous creusons dans le spectre du signal bruité, dans les endroits où l'énergie est faible, et nous respectons le spectre dans les endroits où l'énergie est élevée. La figure 5.3 montre que si l'énergie du signal bruité est élevée, les deux spectres seront confondus sinon, le spectre du signal débruité est plus bas que le spectre du signal bruité.

La figure 5.4 présente deux parties des signaux de la figure 5.1. Ces parties sont des trames de bruit de bafouillage (babble noise). L'énergie de ce bruit étant élevée, le DAV ne déclenche pas, car il ne fait pas la différence entre la vraie parole et la parole de bafouille.

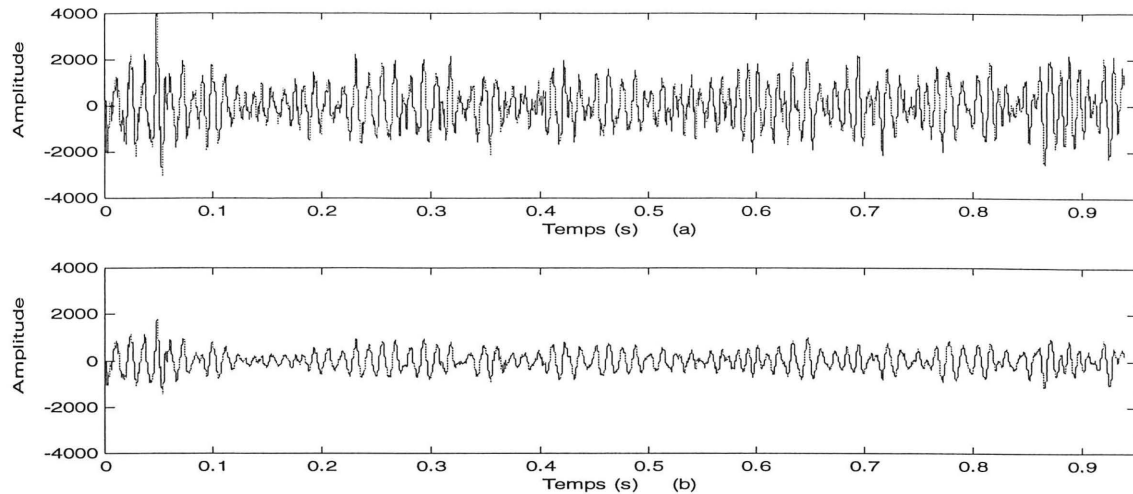


Figure 5.4 Réduction de bruit. (a) Bruit de bafouille avec un RSB de 5 dB. (b) Bruit traité par l'algorithme proposé pour la bande élargie.

Dans le cas d'une présence du bruit sans parole, nous traitons le spectre d'amplitude du bruit différemment. Nous décalons le spectre d'amplitude vers le bas d'un niveau dépendant de l'estimation de l'énergie du bruit calculée. La figure 5.5 décrit les spectres d'amplitude des signaux de la figure 5.4.

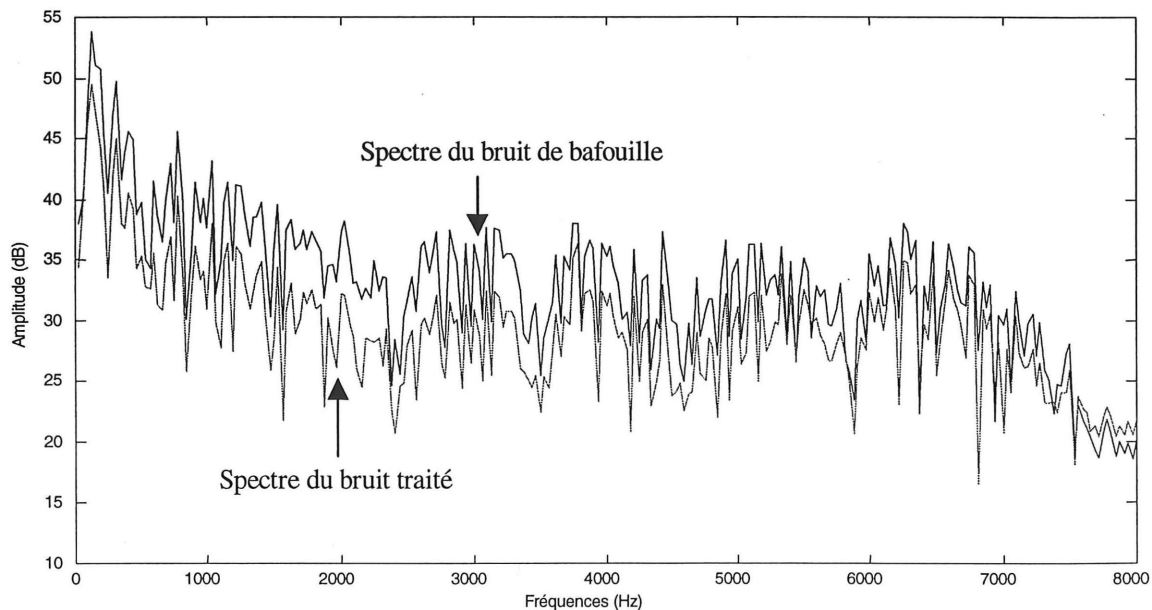


Figure 5.5 Spectres d'amplitude des deux signaux de la figure 5.4.

Le spectre du bruit traité est décalé vers le bas par rapport à celui du spectre du bruit avant traitement. Nous remarquons que le spectre n'est pas plat, car le bruit en question est un bruit de bafouille (babble noise).

Nous avons vu le bruit de bafouille qui n'est pas évident à traiter et les spectres de ces bruits. La figure 5.6 montre un bruit des véhicules et de la rue (vent, bruit variable, ...). Dans ce cas, le DAV déclenche plus efficacement que dans le cas du bruit de bafouille.

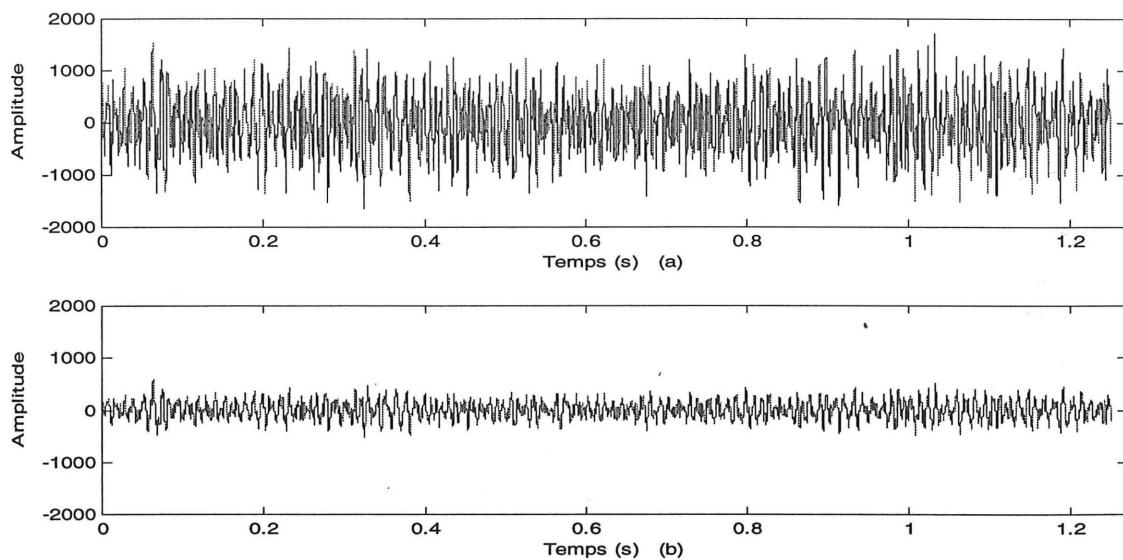


Figure 5.6 Réduction de bruit. (a) Bruit des voitures et de la rue avec un RSB de 10 dB. (b) Bruit traité.

À la figure 5.7, nous montrons les spectres d'amplitudes des signaux de la figure 5.6.

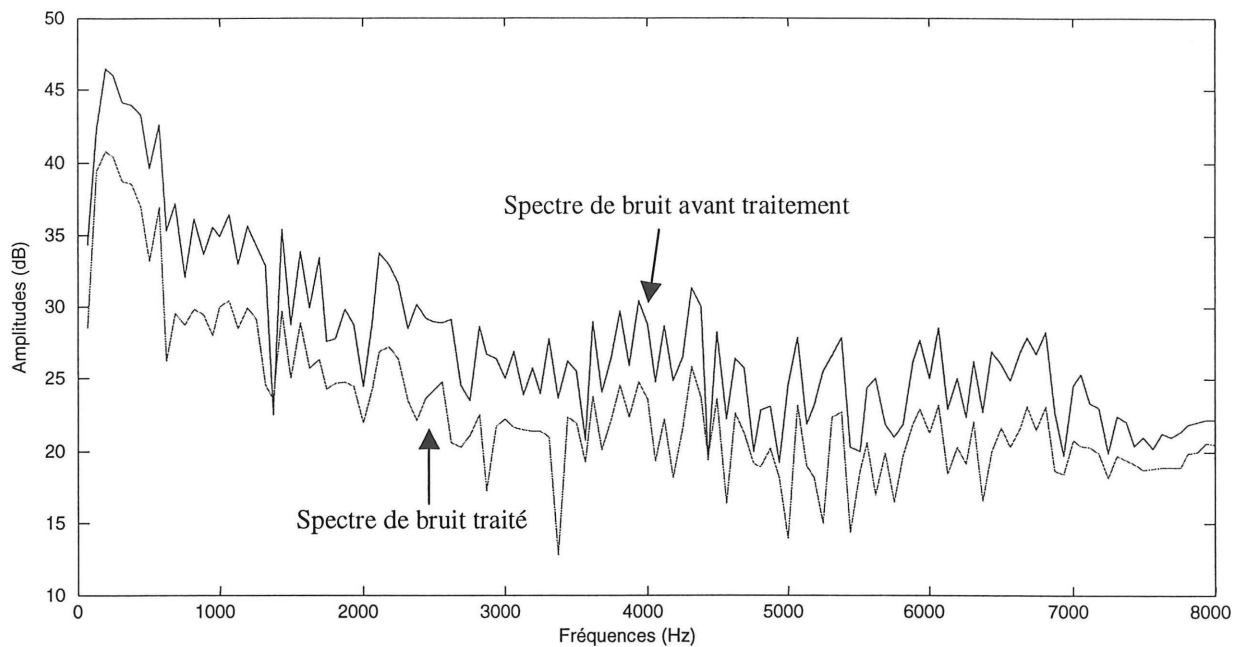


Figure 5.7 Spectres d'amplitude des deux signaux de la figure 5.6.

Les spectres de la figure 5.7 sont plus plats que ceux de la figure 5.5, car le bruit dans ce cas est plus stationnaire.

Finalement nous illustrons dans la figure 5.8 un signal propre (sans bruit) et le signal traité par l'algorithme proposé. Dans le cas d'un signal propre, l'algorithme ne doit pas le changer.

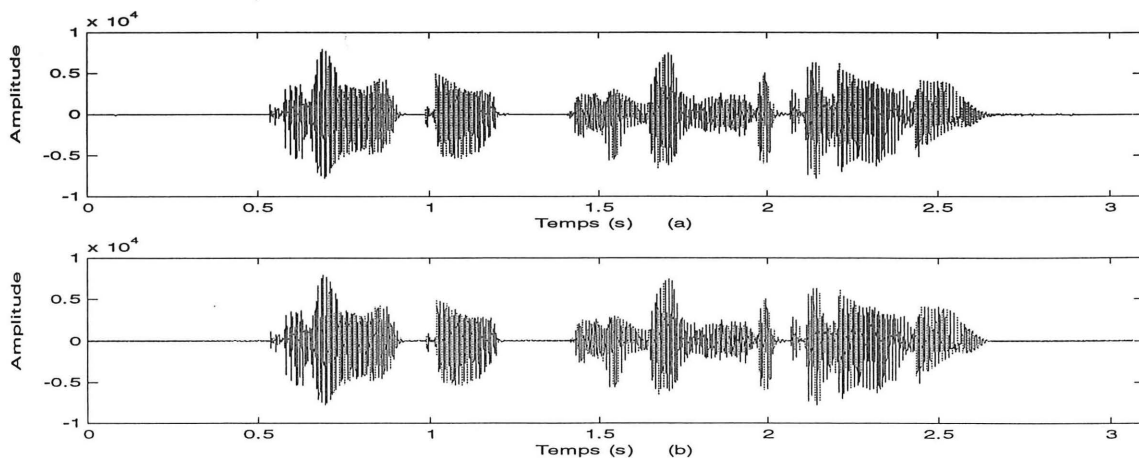


Figure 5.8 (a) Signal de parole propre. (b) Signal traité.

Pour voir si le signal de la figure 5.8 (b) a été modifié ou non, nous présentons les spectres d'amplitudes de ces deux signaux.

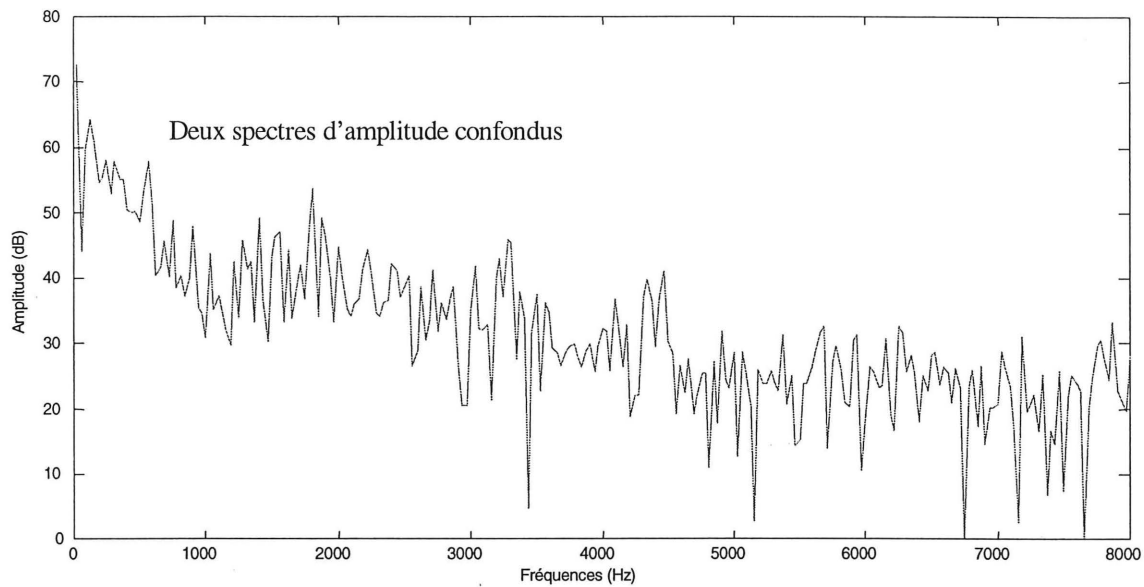


Figure 5.9 Spectres d'amplitudes des deux signaux de la figure 5.8.

Les spectres de la figure 5.9 sont confondus. C'est-à-dire que le signal traité par l'algorithme proposé n'a pas été changé par l'algorithme proposé.

Pour rentrer plus en détails, nous présentons deux parties voisées propres de la figure 5.9. La figure 5.10 les montrent.

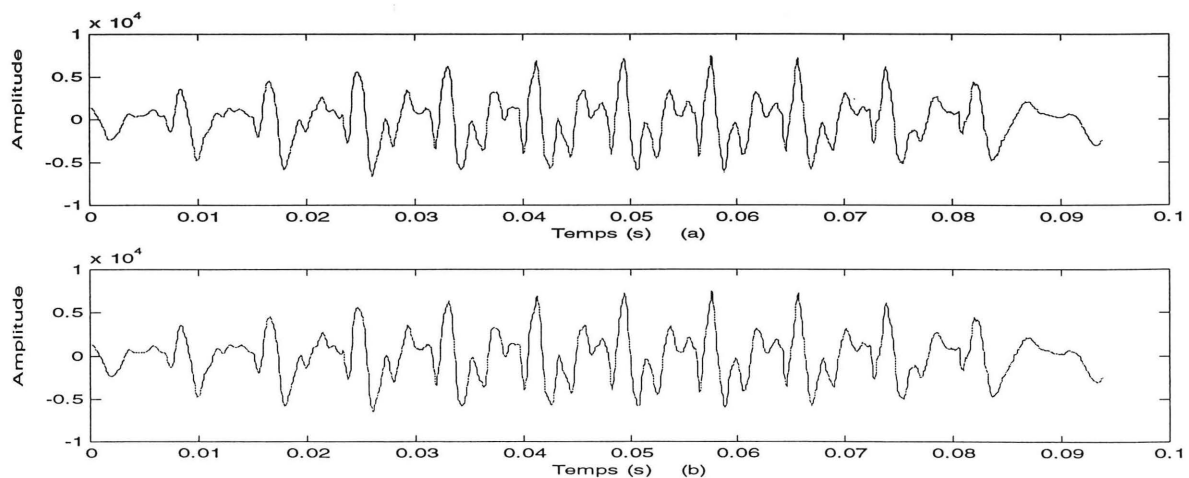


Figure 5.10 (a) Signal propre de parole voisée. (b) Signal traité.

Ensuite les spectres des signaux de la figure 5.11 sont illustrés à la figure 5.12.

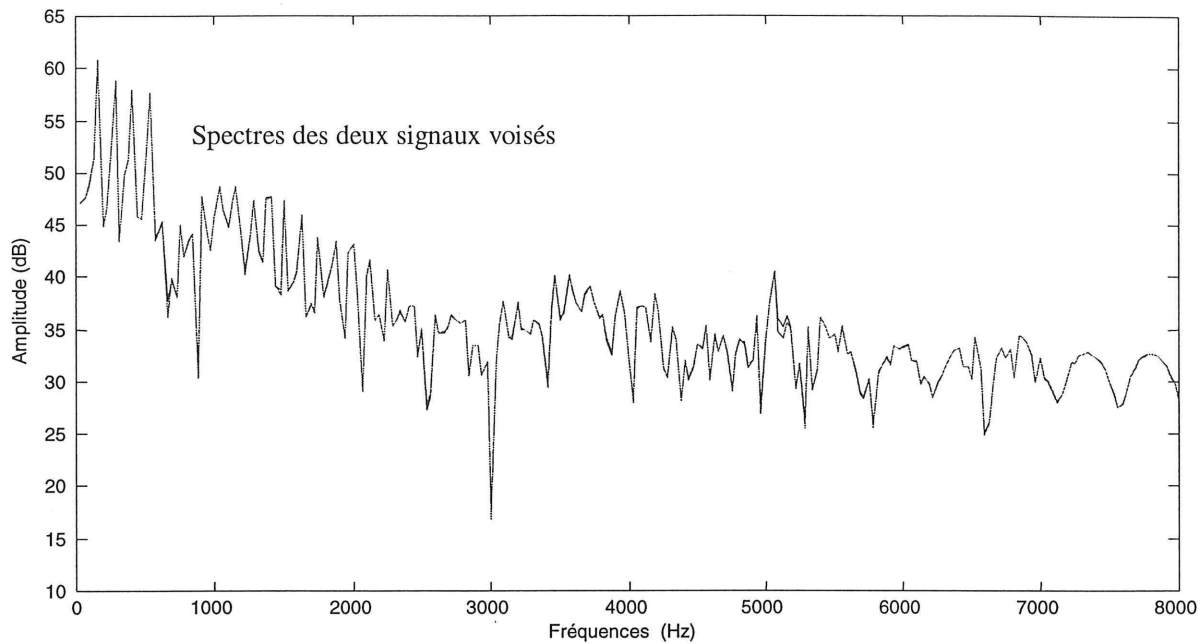


Figure 5.11 Deux Spectres d'amplitude confondus pour les signaux de la figure 5.10.

Nous remarquons sur la figure 5.11 que les deux spectres sont confondus. Donc, l'algorithme proposé n'a pas changé le spectre du signal voisé.

La figure 5.12 décrit deux parties non voisées propres de la figure 5.8.

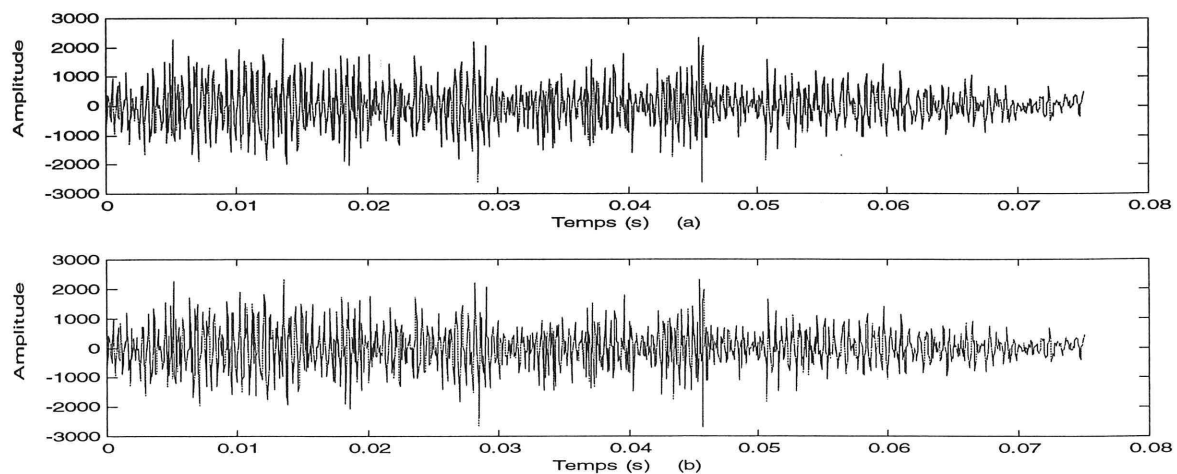


Figure 5.12 (a) Partie non voisée du signal de parole de la figure 5.8 (a).
(b) Signal traité.

Pour ressortir les changements dans le spectre d'amplitude d'un signal non voisé après traitement, nous présentons les spectres suivants :

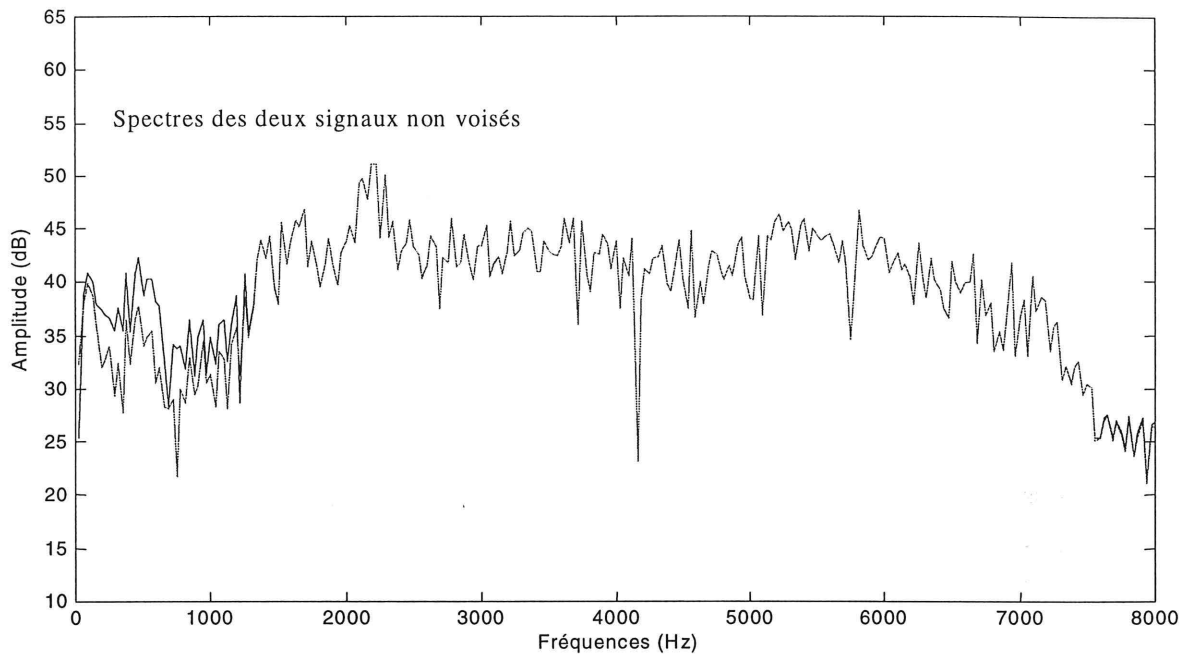


Figure 5.13 Spectres d'amplitude des deux signaux de la figure 5.12.

Dans un signal non voisé, il y a plus d'énergie dans les hautes fréquences que dans les basses; dans ce cas-ci, l'énergie des basses fréquences est faible. Ainsi, pour cela nous remarquons un petit changement dans le spectre d'amplitude du signal débruité à la figure 5.13; ce changement est causé par la présence d'une énergie faible aux basses fréquences. Par contre, aux hautes fréquences les spectres sont confondus. De plus, il peut y avoir un déclenchement du « flag » dans le cas d'un signal non voisé, car le DAV voit des trames non voisés comme du bruit.

La section 5.12.1 présente des statistiques faites lors d'un test par 15 personnes pour évaluer l'algorithme proposé.

Ce test est appelé CCR test (Comparative Category Rating) et la moyenne des votes est appelée CMOS (Comparative Mean Opinion Score).

5.12.1 Statistiques

Nous avons préparé 6 signaux de parole échantillonnés à 16 kHz dont 3 sont des voix d'hommes et 3 celles des voix de femmes. Ensuite, chaque signal propre est bruité avec du bruit différent à deux niveaux (5 dB et 10 dB). Ceux-ci nous donnent 12 signaux de parole dont 6 ont un RSB de 5 dB et les autres ont un RSB de 10 dB. En plus, nous avons fait un test sur 6 signaux de parole propre (sans bruit). Le but de passer un test sur des signaux propres est de voir si l'algorithme proposé affecte un signal de parole propre.

Les exemples de bruit que nous avons testés sont : le bruit de bafouille (babble noise), le bruit de la rue (street noise), le bruit des voitures (car noise) et le bruit du bureau (office noise).

Le test consiste à comparer deux signaux bruités et débruités. Les auditeurs votent tout le temps pour le deuxième signal. L'ordre des signaux bruité et débruité n'est pas respecté, il est tout simplement aléatoire.

Les personnes qui ont fait le test ont voté dans un tableau comme suit :

Tableau 5-2 Notations de l'écoute

Groupe I	
Beaucoup mieux	2
Mieux	1
Égale	0
Mauvais	-1
Très mauvais	-2

La méthode du test CCR mentionne que, pour évaluer le test, nous devons donner aux évaluations les nombres -2, -1, 0, 1 et 2 pour très mauvais, mauvais, égale, mieux et beaucoup mieux respectivement.

Les résultats du test pour les signaux bruités et débruités présente un niveau classé entre mieux et beaucoup mieux. Car la moyenne que nous avons eu était de 1.4. Le nombre que nous avons donné à « mieux » était 1 et à « beaucoup mieux » 2. Alors, 1.4 est un chiffre plus élevé que mieux, ce qui est satisfaisant comme résultat.

Finalement, tous les tests d'écoute sur les signaux propres ont montré une égalité par unanimité avec le signal original; c'est-à-dire que l'algorithme proposé n'a pas modifié le signal propre original. Ainsi, le but a été atteint.

Conclusion

Le bruit de l'environnement dégrade de façon significative le signal de parole. Cette dégradation du signal affecte les performances de la plupart des systèmes de communication par voix.

Plusieurs méthodes ont été introduites pour réduire le niveau du bruit dans les signaux à bande étroite. Une de ces méthodes a été présentée par Ramabadran et al. Mais, cet algorithme n'était pas efficace dans toutes les situations. Pour cela, nous l'avons analysé et modifié pour l'améliorer. Pour réaliser ce travail, nous avons développé un nouveau code avec les changements que nous avons jugés utiles.

Après avoir réalisé le code pour les signaux à bande étroite, nous l'avons étendu pour les signaux à bande élargie. Ce travail n'était pas dans la littérature mais, nous l'avons réalisé suite à notre compréhension de l'algorithme à bande étroite.

Donc, à la fin de ce travail de recherche, nous avons réalisé deux algorithmes qui réduisent, de façon remarquable, le bruit additif au signal de parole propre.

Le premier algorithme est un réducteur de bruit pour la bande téléphonique de 4 kHz. Cet algorithme se base sur la soustraction spectrale et le rapport signal sur bruit (RSB). Le calcul du RSB a pour but d'éviter la soustraction du bruit dans les bandes où le RSB est élevé. Donc, la quantité de bruit à soustraire dépend du niveau du RSB et son emplacement dans les bandes

critiques. Cette façon de faire annule le bruit musical causé par la soustraction spectrale. Les résultats de cet algorithme ont été comparés avec ceux d'un autre algorithme similaire présenté par Ramabadran et al, et se sont avérés meilleurs. Cet algorithme et les résultats qui en découlent ont constitué la base initiale qui a été élargie pour une bande de 8 kHz. Des changements mineurs à l'algorithme à bande étroite ont été apportés pour permettre l'extension à la nouvelle bande.

Pour avancer et améliorer ce travail, nous suggérons de s'attarder sur le déclenchement du DAV, avec l'algorithme de la bande étroite, pour des bruits spécifiques. La réduction du bruit nécessite des tests plus avancés pour optimiser son application.

Bibliographie

- [1] S. F. Boll, « Suppression of Acoustic Noise in Speech Using Spectral Subtraction, » *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, p. 113-120, Avril 1979.
- [2] J. S. Lim et A. V. Oppenheim, « Enhancement of speech bandwidth compression of noise speech, » *Proc. IEEE*, vol. 67, p. 1586-1604, Décembre 1979.
- [3] P. Lockwood et J. Boudy, « Experiments with a nonlinear spectral subtractor (NNS), hidden Markov models and projection, for robust recognition in cars, » *Speech communication*, vol. 11, p. 215-228, Juin 1992.
- [4] M. Berouti, R. Schwartz et J. Makhoul, « Enhancement of speech corrupted by acoustic noise, » *IEEE ICASSP*, Washington, DC, p. 208-211, Avril 1979.
- [5] N. Virag, « Single Channel Speech Enhancement Based On Masking Properties Of the Human Auditory System, » *IEEE Trans. Speech and Audio Processing*, vol. 7, No. 2, p. 126-137, Mars 1999.
- [6] Y. Ephraim et D. Malah, « Speech Enhancement using a minimum mean-square error short-time spectral amplitude estimator, » *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, p.1109-1121, Décembre 1984.
- [7] D. Tsoukalas, M. Praskevas et J. Mourjopoulos, « Speech Enhancement using psycho-acoustic criteria, » *Proc. IEEE ICASSP*, Minneapolis, MN, p. 359-361, Avril 1994.
- [8] T. Usagawa, M. Iwata et m. Ebata, « Speech parameter extracting in noisy environment using a masking model, » *Proc. IEEE ICASSP*, Adelaide, Australia, vol. II, p. 81-84, Avril 1994.

- [9] J. F. Lynch, J. G. Jossenhans, et R. E. Crochiere, « Speech/Silence segmentation for real-time coding via rule based adaptive endpoint detection, » *Proc. IEEE ICASSP*, Dallas, TX, vol. p. 1348-1351, Avril 1987.
- [10] J. D. Johnston, « Transform coding of audio signal using perceptuel noise criteria, » *IEEE J. on select, Areas commun.*, vol. 6, p. 314-323, Février 1988.
- [11] T. V. Ramabadran, J. P. Ashley, M. J. McLaughlin, « Background Noise Suppression For Speech Enhancement And Coding, » *IEEE Speech Coding Workshop*, p. 43-44, 1997.
- [12] H. Drucker, « Speech processing in a high ambient noise environment, » *IEEE Trans. Audio and Electronics*, Vol. AU-16, p. 165-168, Juin 1968.
- [13] R. J. McAulay et M. L. Malpass, « Speech enhancement using a soft-decision noise suppression filter, » *IEEE Trans. Acoustics, Speech, signal Process.*, Vol. 28, p. 137-145, Avril 1980.
- [14] J. S. Lim et A. V. Oppenheim, « All-pole modeling of degraded speech, » *IEEE Trans. Acoustics, Speech, signal Process.*, Vol. 26, p. 197-210, Juin 1978.
- [15] Y. Ephraim, « Statistical-Model-Based Speech Enhancement Systems, » *Proc. IEEE*, Vol. 80, No. 10, Octobre 1992.
- [16] L. Arslan, A. McCree, V. Viswanathan, « New Methods for Adaptive noise suppression, » *IEEE Int. Conf., Acoust., Speech, signal Process. (ICASSP-95)*, 812-815, 1995.
- [17] Y. Ephraim et H. L. Van Trees, « A Signal Subspace Approach for Speech Enhacement, » *IEEE Transaction, Speech And Audio Processing*. Vol. 3, No. 4, p. 251-266, juillet 1995.

- [18] A. Vähätalo et I. Johansson, « Voice Activity Detection for GSM Adaptive Multi-Rate CODEC, » Nokia Research Centre, P.O.Box 100, FIN-33721 Tampere, Finland et Ericsson Erisoft AB, SE-971 28 Lulea, Sweden.
- [19] M. Jelinek et F. Labonté, « Robust Signal/Noise Discrimination for Wideband Speech and Audio Coding, » sera publié au *IEEE Speech Coding Workshop*, 2000.
- [20] GSM recommendation for VAD : GSM 06.32, GSM 06.42 et GSM 06.82.
- [21] ITU-T, Rec G.723.1 Annexe A : « Silence Compression Schema ».
- [22] ITU-T, Rec G.729 Annexe B: « A Silence Compression Schema for G.729 Optimized for terminals conforming to ITU-T v.70 ».

Annexe A

Liste des symboles et abréviations

DFT	Transformation de Fourier Discrète
IDFT	Transformation de Fourier Discrète Inverse
E_b	Énergie du Bruit
E_c	Énergie par bande
N_b	Nombre des bandes
RSB	Rapport Signal sur Bruit
DAV	Détection d'Activité Vocale
γ_c	Gain par bande
γ_g	Gain général par bande
γ_{dB}	Gain en décibel par bande